electronic

networking:

research, applications and policy

Volume 2, Number 1 · Spring 1992

A Special Issue of Electronic Networking: Research, Applications, and Policy

Accessing Information on the Internet

a Meckler publication

EDITORIAL ADDRESSES

EDITORS

Charles R. McClure, Editor
School of Information Studies
Center for Science and Technology
Rm. 4-206
Syracuse University
Syracuse NY 13244
Phone: 315-443-2911
Fax: 315-443-5806
(cmcclure@suvm.acs.syr.edu)

Ann P. Bishop, Associate Editor
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
426 David Kinley Hall
1407 W. Gregory Drive
Urbana, IL 61801
Phone: 217-244-3299
Fax: 217-244-3302
(abishop@uiuc.edu)

Philip Doty, Associate Editor
Graduate School of Library and Information Science
University of Texas at Austin
Austin, TX 78712-1276
Phone: 512-471-3839
Fax: 512-471-3971
(pdoty@utxvm.cc.utexas.edu)

Joe Ryan, Resource Review Editor
School of Information Studies
Center for Science and Technology
Rm. 4-206
Syracuse University
Syracuse NY 13244
Phone: 315-443-2911
Fax: 315-443-5806
(joryan@suvm.acs.syr.edu)

MEMBERS

Pamela Q. J. Andre
Associate Director for Automation
National Agricultural Library
10301 Baltimore Boulevard
Beltsville, MD 20705
Phone: 301 344-3813
Fax: 301 344-5473
Internet: pandre@asrr.arsusda.gov

Michael Breaks
University Librarian
Heriot-Watt University
Riccarton
Edinburgh EH14 4AS
Scotland
Phone: 031-449-5111
Fax: 031-449-5153
Internet: libmlb@vaxb.hw.ac.uk

Martin Dillon, Ph. D.
Director
Office of Research
OCLC
6565 Frantz Road
Dublin, OH 43017-0702
Phone: 614-764-4365
Fax: 614 764-6096
Internet: mjd@rsch.oclc.org

Susan Estrada
Executive Director, CERFnet
General Atomics
P.O. Box 85608
San Diego, CA 92186-9784
Phone: 619-534-5067
Fax: 619-534-5167
Internet: estradas@sdsc.edu

Brian Kahin, Esq.
Director
Information Infrastructure Project
Science, Technology and Public Policy Program
John F. Kennedy School of Government
Harvard University
Cambridge, MA 02138
Phone: 617-495-8903
Fax: 617-495-5776
Internet: kahin@hulaw1.harvard.edu

Michael J. McGill, Ph.D.
Director of Network Systems
Information Technology Division
5112 Institute for Science and Technology
2900 Bonistell
University of Michigan
Ann Arbor, MI 48109-2099
Internet: mcgill@oar.net
Phone: 313-764-5442

Pat Molholt, Associate Director
Institute Libraries
Rensselaer Polytechnic Institute
Troy, NY 12180 - 3590
Phone: 518-276-8300
Fax: 518-276-8559
Internet: usera0pm@rpitsmts

Tracy LaQuey Parker
Network Information Specialist
Network Services, Computation Center
University of Texas at Austin
Austin, TX 78712
Phone: 512-471-5046
Fax: 512-471-1582
Internet: tracy@utexas.edu

Carol A. Parkhurst
Assistant University Librarian for Systems & Technical Services
University Library/322
University of Nevada, Reno
Reno, NV 89557
Phone: 702-784-6566
Fax: 702-784-1751
Internet: carolp@unr.edu
BITNET: carolp@equinox.bitnet

Henry H. Perritt, Jr.
Professor of Law
Villanova University School of Law
Villanova, PA 19085
Phone: 215-645-7078
Fax: 215-645-7033
Internet: perritt@ucis.vill.edu

Fred W. Weingarten, Ph.D.
Executive Director
Computing Research Association
1625 Massachusetts Ave.
Suite 110
Washington, DC 20036-2212
Phone: (202) 234-2111
Fax: (202) 667-1066
Internet: rweingar@cs.umd.edu

Electronic Networking (ISSN 1051-4805) is published quarterly by Meckler Corporation, 11 Ferry Lane West, Westport, CT 06880 (203) 226-6967; Internet: Meckler@jvnc.net; CompuServe (70373, 616).

Subscriptions: $95.00 (£60.00) per year. Please add $15.00 for subscriptions to Canada and Central and South America. Orders from North, Central, and South America should be sent to *Electronic Networking*, Meckler Corporation, 11 Ferry Lane West, Westport, CT 06880; from elsewhere to Meckler Ltd., 247–249 Vauxhall Bridge Road, London SW1V 1HQ, UK, 01-931-9985. All back issues are available. *Electronic Networking: Research, Applications, and Policy* is indexed by ERIC for the ERIC database, by *Information Science Abstracts*, and also by *Current Journals in Education*.

POSTMASTER: Send all address changes to *Electronic Networking*, Meckler Corporation, 11 Ferry Lane West, Westport, CT 06880.

electronic
networking:
research, applications and policy

Contents

Vol. 2 ▪ No. 1 ▪ Spring 1992

National Research and Education Network (NREN) Budget
(in millions of $)

Agency	FY 1992 (actual)	FY 1993 (requested)
• Defense Advanced Research Projects Agency	32.9	43.6
• National Science Foundation	32.7	45.1
• Department of Energy	12.0	14.0
• National Aeronautics and Space Administration	7.4	9.8
• Department of Health and Human Services	4.2	7.2
• National Oceanic and Atmospheric Administration	0.7	0.4
• Environmental Protection Agency	0	0.4
• National Institute of Standards and Technology	2.0	2.0
Totals	**91.9**	**122.5**

Table 2.

Not mentioned in the law, but having an impact on what will be done and how it will be done, is the FNC which includes participants from the key mission agencies related to HPCC and NREN. Currently, the role of the FNC and its advisory committee is not well understood or publicized. The FNC and its advisory committee must do a better job of informing interested individuals of (1) the Council's activities, (2) the Council's recommendations, and (3) how public input to the Council and its advisory committee can be best channeled.

One wonders how the role of the FNC might be affected by section 101 (b) which establishes the "High Performance Computing Advisory Committee." Some description of each agency's responsibilities can be found in Title II of the law, but it is unclear how the advisory committee would affect agency decision making. The advice and recommendations of the FNC and the Advisory Committee, however, may have a significant impact on the programs, funding levels, and specific initiatives likely to be taken when implementing P.L. 102-194.

The range of goals espoused for both HPCC and the NREN will require setting priorities and determining which program initiatives are most important. Given the disparity between the proposed budgets for HPCC and the NREN, one wonders how federal policymakers will allocate resources for NREN applications, services, and education and training. Will these areas fall through the budgetary cracks?

How the various stakeholders can best affect policy development at the FNC, at OSTP, and at individual agencies is a matter of some concern. This diffuse and decentralized policy area will not only be difficult to manage and coordinate, but it may be difficult for citizens and public advocacy groups to make their views known and have them taken seriously.

Access to the NREN

Section 102 (b) offers some interesting language related to access to the NREN and deserves a very careful reading. Seen by some observers as one of the most important sections of the proposed NREN program, it addresses the need for and importance of access to the NREN by a range of stakeholders. But the sense of this section is severely compromised by phrases such as "as appropriate," "with appropriate," and "to the extent possible."

This section is a good example of what one federal policymaker told this writer was "jello" language—meaning that the issues and implementation strategies will be debated and developed "later." To a large degree, however, the language in this section provides a carte blanche for federal policymakers to develop policies on access that could range from extremely restrictive to extremely accessible. Interested individuals and stakeholder groups will need to lobby policymakers about what "appropriate access" and to the "extent possible" to electronic information resources really means.

The Missing Department of Education

How does the Department of Education fit into P.L. 102-194? Section 206 outlines responsibilities for the department as "being authorized to conduct basic and applied research in computational research with an emphasis on the coordination of activities with libraries, school facilities and education research groups." Despite heroic efforts on the part of a number of public advocacy groups, the department's role is limited. Indeed, were it not for these lobbying efforts, there may not have been any role!

But the proposed authorizations for NREN initiatives from the Department of Education in P.L. 102-194 are minute—$1.7 million for FY 1993. Compared to the proposed budget requests for other agencies, as outlined above, $1.7 million will result in the Department of Education having little presence in the NREN initiatives.

Moreover, the Department of Education is in the middle of its own initiative, *America 2000: An Education Strategy* (1991). One of their initiatives is "Bringing America On-Line." But no mention of bringing American on-line via the NREN is made in *America 2000*. In fact, there appears to be limited coordination and joint planning between the NREN initiatives and *America 2000*. Specific language in P.L. 102-194 regarding educational initiatives, bringing in the K-12 audience, and linking the library and educational community into the NREN are conspicuous by the limited attention they received.

Dissemination of Government Information

One of the most interesting portions of the new law is Section 101(2) (E) which states that the HPCP shall "provide for improved dissemination of Federal agency data and electronic information." This seemingly simple statement belies an exceedingly complex federal information policy system.

The current decentralized and ambiguous Federal information policy system is based primarily on agency-specific statutes regarding the dissemination of government information, for example, 44 *U.S.C.*, dealing with the Government Printing Office; more general statutes such as the Paperwork Reduction Act, Copyright, and Privacy Acts; and regulations such as the Office of Management and Budget (OMB) Circular A-130, "The Management of Federal Information Resources" (Hernon & McClure, 1987).

A number of new initiatives are also being considered regarding dissemination of government information in electronic format. For example:

- Reauthorization of the Paperwork Reduction of Act of 1991 [S. 1044, Glenn Bill; S. 1139, Nunn Bill]
- Government Printing Office Wide Information Network for Data Online Act of 1991 (WINDO) [H.R. 2772]
- Improvement of Information Access Act [H.R. 3459]
- Revision of OMB Circular A-130, "The Management of Federal Information Resources."

Each of these, and other proposals not mentioned here, deal with improving the dissemination of government information in, or through the use of, electronic formats. How HPCC and the NREN will interface with existing electronic information dissemination policy, as well as recent initiatives such as those above, will require careful analysis and much debate.

Next Steps

The issues discussed in this brief editorial only scratch the surface of policy areas that will require additional public debate and discussion. Other key issues to be addressed relate to pricing of NREN services, commercialization of networked information, insuring equitable access to information resources for all members of society, supporting testbed institutions to develop new network technologies, strategies for network education and training, protecting intellectual property over the network, and a host of others.

Some individuals and public advocacy groups believe that additional legislation may be needed as a followup to the High Performance Computing Act of 1991 to address questions related to management and coordination of HPCC and the NREN; to better describe specific program initiatives rather than allowing individual agency determination of what should be done and how; and to better define the role and involvement of the library and education community in the NREN.

But rather than seek additional legislation, interested stakeholder groups should become knowledgeable about the issues, and work within the existing policy system to make their views known to affect policy development of issues such as those discussed here and to be active at the agency level in putting forth ideas and strategies for implementation. An excellent example of this kind of work is

that being done by the Coalition for Networked Information.

A critical first step for stakeholders interested in the development of HPCC and the NREN is to carefully read the legislation (see appendix), review *Grand Challenges 1992: High Performance Computing and Communications* (1992), familiarize themselves with supporting background materials produced in the policy debates (McClure, Bishop, Doty, & Rosenbaum, 1991), reach their own conclusions as to what policy initiatives should be developed, and make their views known to appropriate public advocacy groups and federal policymakers. Some of the many places to make those views known are:

- Your congressional representatives and senators
- Office of Science and Technology Policy, New Executive Office Building, Washington DC, 20506
- Committee on Science, Commerce, and Transportation, Hart Building, Suite 427,U.S. Senate, Washington, DC, 20510
- Chairperson, Federal Networking Council, National Science Foundation, 1800 G Street, NW, Washington, DC, 20550
- Coalition for Networked Information, 1527 New Hampshire Ave., NW, Washington, DC, 20036.

Because OSTP must issue a report, one year after passage of P.L. 102–104, on the status of HPCC and NREN [see Section 102(g)], it is especially important to direct comments to this agency. The six issue areas that OSTP must address are funding, fees, future operations, copyright protection, commercial traffic, and security.

You may also wish to initiate a discussion of issues within your own institutions or professional groups. Moreover, we welcome your letters to *Electronic Networking: Research, Applications, and Policy* regarding the next steps for developing the HPCC and NREN programs.

The point is to make your views known and get involved. The successful development of the HPCP and the NREN is dependent on ongoing open and active policy debates among *informed* individuals. The issues are of critical importance; ideas, debate, proposals, and strategy development are needed now. The passage of the High Performance Computing Act of 1991 into P.L. 102-194 is a beginning, not an end.

Note

1. Copies of *Grand Challenges 1993: High Performance Computing and Communications* can be obtained from: Federal Coordinating Council for Science, Engineering, and Technology. Committee on Physical, Mathematical, and Engineering Sciences, c/o the National Science Foundation, Computer and Information Science and Engineering Directorate, 1800 G Street, NW, Washington, DC 20550.

References

America 2000: An education strategy, sourcebook. (1991). Washington, DC: U.S. Department of Education [ED/OS91-13].

Executive Office of the President. (1992). Budget of the United States government fiscal year 1993. Washington, DC: Government Printing Office.

Grand challenges 1993: High performance computing and communications, The FY 1993 U.S. research and development program [to supplement the President's Fiscal Year 1993 Budget]. (1992). Washington, DC: Committee on Physical, Mathematical, and Engineering Sciences. Federal Coordinating Council for Science, Engineering, and Technology: Office of Science and Technology Policy.

Hernon, Peter, & McClure, Charles R. (1987). *Federal information policies in the 1980s.* Norwood, NJ: Ablex Publishing Corporation.

McClure, Charles R., Bishop, Ann, Doty, Philip, & Rosenbaum, Howard. (1991). *The National Research and Education Network: Research and policy prspectives.* Norwood, NJ: Ablex Publishing Corporation, 1991 [appendices contain the original bills, hearings, and other key source documents].

Appendix

High-Performance Computing Act of 1991

Public Law 102-194
102d Congress

An Act

Dec. 9, 1991
[S. 272]

To provide for a coordinated Federal program to ensure continued United States leadership in high-performance computing.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

High-Performance Computing Act of 1991. 15 USC 5501 note.

SECTION 1. SHORT TITLE.

This Act may be cited as the "High-Performance Computing Act of 1991".

15 USC 5501.

SEC. 2. FINDINGS.

The Congress finds the following:

(1) Advances in computer science and technology are vital to the Nation's prosperity, national and economic security, industrial production, engineering, and scientific advancement.

(2) The United States currently leads the world in the development and use of high-performance computing for national security, industrial productivity, science, and engineering, but that lead is being challenged by foreign competitors.

(3) Further research and development, expanded educational programs, improved computer research networks, and more effective technology transfer from government to industry are necessary for the United States to reap fully the benefits of high-performance computing.

(4) A high-capacity and high-speed national research and education computer network would provide researchers and educators with access to computer and information resources and act as a test bed for further research and development of high-capacity and high-speed computer networks.

(5) Several Federal agencies have ongoing high-performance computing programs, but improved long-term interagency coordination, cooperation, and planning would enhance the effectiveness of these programs.

(6) A 1991 report entitled "Grand Challenges: High-Performance Computing and Communications" by the Office of Science and Technology Policy, outlining a research and development strategy for high-performance computing, provides a framework for a multiagency high-performance computing program. Such a program would provide American researchers and educators with the computer and information resources they need, and demonstrate how advanced computers, high-capacity and high-speed networks, and electronic data bases can improve the national information infrastructure for use by all Americans.

5 USC 5502.

SEC. 3. PURPOSE.

The purpose of this Act is to help ensure the continued leadership of the United States in high-performance computing and its applications by—

(1) expanding Federal support for research, development, and application of high-performance computing in order to—

(A) establish a high-capacity and high-speed National Research and Education Network;

(B) expand the number of researchers, educators, and students with training in high-performance computing and access to high-performance computing resources;

(C) promote the further development of an information infrastructure of data bases, services, access mechanisms, and research facilities available for use through the Network;

(D) stimulate research on software technology;

(E) promote the more rapid development and wider distribution of computing software tools and applications software;

(F) accelerate the development of computing systems and subsystems;

(G) provide for the application of high-performance computing to Grand Challenges;

(H) invest in basic research and education, and promote the inclusion of high-performance computing into educational institutions at all levels; and

(I) promote greater collaboration among government, Federal laboratories, industry, high-performance computing centers, and universities; and

(2) improving the interagency planning and coordination of Federal research and development on high-performance computing and maximizing the effectiveness of the Federal Government's high-performance computing efforts.

15 USC 5503.

SEC. 4. DEFINITIONS.

As used in this Act, the term—

(1) "Director" means the Director of the Office of Science and Technology Policy;

(2) "Grand Challenge" means a fundamental problem in science or engineering, with broad economic and scientific impact, whose solution will require the application of high-performance computing resources;

(3) "high-performance computing" means advanced computing, communications, and information technologies, including scientific workstations, supercomputer systems (including vector supercomputers and large scale parallel systems), high-capacity and high-speed networks, special purpose and experimental systems, and applications and systems software;

(4) "Network" means a computer network referred to as the National Research and Education Network established under section 102; and

(5) "Program" means the National High-Performance Computing Program described in section 101.

TITLE I—HIGH-PERFORMANCE COMPUTING AND THE NATIONAL RESEARCH AND EDUCATION NETWORK

15 USC 5511.

SEC. 101. NATIONAL HIGH-PERFORMANCE COMPUTING PROGRAM.

President.

(a) National High-Performance Computing Program.—(1) The President shall implement a National High-Performance Computing Program, which shall—

(A) establish the goals and priorities for Federal high-performance computing research, development, networking, and other activities; and

(B) provide for interagency coordination of Federal high-performance computing research, development, networking, and other activities undertaken pursuant to the Program.

(2) The Program shall—

(A) provide for the establishment of policies for management and access to the Network;

(B) provide for oversight of the operation and evolution of the Network;

(C) promote connectivity among computer networks of Federal agencies and departments;

(D) provide for efforts to increase software availability, productivity, capability, portability, and reliability;

(E) provide for improved dissemination of Federal agency data and electronic information;

(F) provide for acceleration of the development of high-performance computing systems, subsystems, and associated software;

(G) provide for the technical support and research and development of high-performance computing software and hardware needed to address Grand Challenges;

(H) provide for educating and training additional undergraduate and graduate students in software engineering, computer science, library and information science, and computational science; and

(I) provide—

(i) for the security requirements, policies, and standards necessary to protect Federal research computer networks and information resources accessible through Federal research computer networks, including research required to establish security standards for high-performance computing systems and networks; and

(ii) that agencies and departments identified in the annual report submitted under paragraph (3)(A) shall define and implement a security plan consistent with the Program and with applicable law.

(3) The Director shall—

(A) submit to the Congress an annual report, along with the President's annual budget request, describing the implementation of the Program; Reports.

(B) provide for interagency coordination of the Program; and

(C) consult with academic, State, industry, and other appropriate groups conducting research on and using high-performance computing.

(4) The annual report submitted under paragraph (3)(A) shall—

(A) include a detailed description of the goals and priorities established by the President for the Program;

(B) set forth the relevant programs and activities, for the fiscal year with respect to which the budget submission applies, of each Federal agency and department, including—

(i) the Department of Agriculture;
(ii) the Department of Commerce;
(iii) the Department of Defense;
(iv) the Department of Education;
(v) the Department of Energy;
(vi) the Department of Health and Human Services;
(vii) the Department of the Interior;
(viii) the Environmental Protection Agency;
(ix) the National Aeronautics and Space Administration;
(x) the National Science Foundation; and
(xi) such other agencies and departments as the President or the Director considers appropriate;

(C) describe the levels of Federal funding for the fiscal year during which such report is submitted, and the levels proposed for the fiscal year with respect to which the budget submission applies, for specific activities, including education, research, hardware and software development, and support for the establishment of the Network;

(D) describe the levels of Federal funding for each agency and department participating in the Program for the fiscal year during which such report is submitted, and the levels proposed for the fiscal year with respect to which the budget submission applies; and

(E) include an analysis of the progress made toward achieving the goals and priorities established for the Program.

(b) High-Performance Computing Advisory Committee.—The President shall establish an advisory committee on high-performance computing consisting of non-Federal members, including representatives of the research, education, and library communities, network providers, and industry, who are specially qualified to provide the Director with advice and information on high-performance computing. The recommendations of the advisory committee shall be considered in reviewing and revising the Program. The advisory committee shall provide the Director with an independent assessment of— President.

(1) progress made in implementing the Program;

(2) the need to revise the Program;

(3) the balance between the components of the Program;

(4) whether the research and development undertaken pursuant to the Program is helping to maintain United States leadership in computing technology; and

(5) other issues identified by the Director.

(c) Office of Management and Budget.—(1) Each Federal agency and department participating in the Program shall, as part of its annual request for appropriations to the Office of Management and Budget, submit a report to the Office of Management and Budget which— Reports.

(A) identifies each element of its high-performance computing activities which contributes directly to the Program or benefits from the Program; and

(B) states the portion of its request for appropriations that is allocated to each such element.

Appendix

High-Performance Computing Act of 1991 (continued)

(2) The Office of Management and Budget shall review each such report in light of the goals, priorities, and agency and departmental responsibilities set forth in the annual report submitted under subsection (a)(3)(A), and shall include, in the President's annual budget estimate, a statement of the portion of each appropriate agency's or department's annual budget estimate relating to its activities undertaken pursuant to the Program.

15 USC 5512. SEC. 102. NATIONAL RESEARCH AND EDUCATION NETWORK.

(a) ESTABLISHMENT.—As part of the Program, the National Science Foundation, the Department of Defense, the Department of Energy, the Department of Commerce, the National Aeronautics and Space Administration, and other agencies participating in the Program shall support the establishment of the National Research and Education Network, portions of which shall, to the extent technically feasible, be capable of transmitting data at one gigabit per second or greater by 1996. The Network shall provide for the linkage of research institutions and educational institutions, government, and industry in every State.

(b) ACCESS.—Federal agencies and departments shall work with private network service providers, State and local agencies, libraries, educational institutions and organizations, and others, as appropriate, in order to ensure that the researchers, educators, and students have access, as appropriate, to the Network. The Network is to provide users with appropriate access to high-performance computing systems, electronic information resources, other research facilities, and libraries. The Network shall provide access, to the extent practicable, to electronic information resources maintained by libraries, research facilities, publishers, and affiliated organizations.

(c) NETWORK CHARACTERISTICS.—The Network shall—

(1) be developed and deployed with the computer, telecommunications, and information industries;

(2) be designed, developed, and operated in collaboration with potential users in government, industry, and research institutions and educational institutions;

(3) be designed, developed, and operated in a manner which fosters and maintains competition and private sector investment in high-speed data networking within the telecommunications industry;

(4) be designed, developed, and operated in a manner which promotes research and development leading to development of commercial data communications and telecommunications standards, whose development will encourage the establishment of privately operated high-speed commercial networks;

(5) be designed and operated so as to ensure the continued application of laws that provide network and information resources security measures, including those that protect copyright and other intellectual property rights, and those that control access to data bases and protect national security;

(6) have accounting mechanisms which allow users or groups of users to be charged for their usage of copyrighted materials available over the Network and, where appropriate and technically feasible, for their usage of the Network;

(7) ensure the interoperability of Federal and non-Federal computer networks, to the extent appropriate, in a way that allows autonomy for each component network;

(8) be developed by purchasing standard commercial transmission and network services from vendors whenever feasible, and by contracting for customized services when not feasible, in order to minimize Federal investment in network hardware;

(9) support research and development of networking software and hardware; and

(10) serve as a test bed for further research and development of high-capacity and high-speed computing networks and demonstrate how advanced computers, high-capacity and high-speed computing networks, and data bases can improve the national information infrastructure.

(d) DEFENSE ADVANCED RESEARCH PROJECTS AGENCY RESPONSIBILITY.—As part of the Program, the Department of Defense, through the Defense Advanced Research Projects Agency, shall support research and development of advanced fiber optics technology, switches, and protocols needed to develop the Network.

(e) INFORMATION SERVICES.—The Director shall assist the President in coordinating the activities of appropriate agencies and departments to promote the development of information services that could be provided over the Network. These services may include the provision of directories of the users and services on computer networks, data bases of unclassified Federal scientific data, training of users of data bases and computer networks, access to commercial information services for users of the Network, and technology to support computer-based collaboration that allows researchers and educators around the Nation to share information and instrumentation.

(f) USE OF GRANT FUNDS.—All Federal agencies and departments are authorized to allow recipients of Federal research grants to use grant moneys to pay for computer networking expenses.

(g) REPORT TO CONGRESS.—Within one year after the date of enactment of this Act, the Director shall report to the Congress on—

(1) effective mechanisms for providing operating funds for the maintenance and use of the Network, including user fees, industry support, and continued Federal investment;

(2) the future operation and evolution of the Network;

(3) how commercial information service providers could be charged for access to the Network, and how Network users could be charged for such commercial information services;

(4) the technological feasibility of allowing commercial information service providers to use the Network and other federally funded research networks;

(5) how to protect the copyrights of material distributed over the Network; and

(6) appropriate policies to ensure the security of resources available on the Network and to protect the privacy of users of networks.

TITLE II—AGENCY ACTIVITIES

SEC. 201. NATIONAL SCIENCE FOUNDATION ACTIVITIES. 15 USC 5521.

(a) GENERAL RESPONSIBILITIES.—As part of the Program described in title I—

(1) the National Science Foundation shall provide computing and networking infrastructure support for all science and engineering disciplines, and support basic research and human resource development in all aspects of high-performance computing and advanced high-speed computer networking;

(2) to the extent that colleges, universities, and libraries cannot connect to the Network with the assistance of the private sector, the National Science Foundation shall have primary responsibility for assisting colleges, universities, and libraries to connect to the Network;

(3) the National Science Foundation shall serve as the primary source of information on access to and use of the Network; and

(4) the National Science Foundation shall upgrade the National Science Foundation funded network, assist regional networks to upgrade their capabilities, and provide other Federal departments and agencies the opportunity to connect to the National Science Foundation funded network.

(b) AUTHORIZATION OF APPROPRIATIONS.—From sums otherwise authorized to be appropriated, there are authorized to be appropriated to the National Science Foundation for the purposes of the Program $213,000,000 for fiscal year 1992; $262,000,000 for fiscal year 1993; $305,000,000 for fiscal year 1994; $354,000,000 for fiscal year 1995; and $413,000,000 for fiscal year 1996.

SEC. 202. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION ACTIVITIES. 5 USC 5522.

(a) GENERAL RESPONSIBILITIES.—As part of the Program described in title I, the National Aeronautics and Space Administration shall conduct basic and applied research in high-performance computing, particularly in the field of computational science, with emphasis on aerospace sciences, earth and space sciences, and remote exploration and experimentation.

(b) AUTHORIZATION OF APPROPRIATIONS.—From sums otherwise authorized to be appropriated, there are authorized to be appropriated to the National Aeronautics and Space Administration for the purposes of the Program $72,000,000 for fiscal year 1992; $107,000,000 for fiscal year 1993; $134,000,000 for fiscal year 1994; $151,000,000 for fiscal year 1995; and $145,000,000 for fiscal year 1996.

SEC. 203. DEPARTMENT OF ENERGY ACTIVITIES. USC 5523.

(a) GENERAL RESPONSIBILITIES.—As part of the Program described in title I, the Secretary of Energy shall—

(1) perform research and development on, and systems evaluations of, high-performance computing and communications systems;

(2) conduct computational research with emphasis on energy applications;

(3) support basic research, education, and human resources in computational science; and

(4) provide for networking infrastructure support for energy-related mission activities.

(b) COLLABORATIVE CONSORTIA.—In accordance with the Program, the Secretary of Energy shall establish High-Performance Computing Research and Development Collaborative Consortia by soliciting and selecting proposals. Each Collaborative Consortium shall—

(1) conduct research directed at scientific and technical problems whose solutions require the application of high-performance computing and communications resources;

(2) promote the testing and uses of new types of high-performance computing and related software and equipment;

(3) serve as a vehicle for participating vendors of high-performance computing systems to test new ideas and technology in a sophisticated computing environment; and

(4) be led by a Department of Energy national laboratory, and include participants from Federal agencies and departments, researchers, private industry, educational institutions, and others as the Secretary of Energy may deem appropriate.

(c) TECHNOLOGY TRANSFER.—The results of research and development carried out under this section shall be transferred to the private sector and others in accordance with applicable law.

(d) ANNUAL REPORTS TO CONGRESS.—Within one year after the date of enactment of this Act and every year thereafter, the Secretary of Energy shall transmit to the Congress a report on activities taken to carry out this Act.

(e) AUTHORIZATION OF APPROPRIATIONS.—(1) There are authorized to be appropriated to the Secretary of Energy for the purposes of the Program $93,000,000 for fiscal year 1992; $110,000,000 for fiscal year 1993; $138,000,000 for fiscal year 1994; $157,000,000 for fiscal year 1995; and $169,000,000 for fiscal year 1996.

(2) There are authorized to be appropriated to the Secretary of Energy for fiscal years 1992, 1993, 1994, 1995, and 1996, such funds as may be necessary to carry out the activities that are not part of the Program but are authorized by this section.

Appendix

High-Performance Computing Act of 1991 (continued)

15 USC 5524. **SEC. 204. DEPARTMENT OF COMMERCE ACTIVITIES.**

(a) GENERAL RESPONSIBILITIES.—As part of the Program described in title I—

(1) the National Institute of Standards and Technology shall—

(A) conduct basic and applied measurement research needed to support various high-performance computing systems and networks;

(B) develop and propose standards and guidelines, and develop measurement techniques and test methods, for the interoperability of high-performance computing systems in networks and for common user interfaces to systems; and

(C) be responsible for developing benchmark tests and standards for high-performance computing systems and software; and

(2) the National Oceanic and Atmospheric Administration shall conduct basic and applied research in weather prediction and ocean sciences, particularly in development of new forecast models, in computational fluid dynamics, and in the incorporation of evolving computer architectures and networks into the systems that carry out agency missions.

(b) HIGH-PERFORMANCE COMPUTING AND NETWORK SECURITY.—Pursuant to the Computer Security Act of 1987 (Public Law 100-235; 101 Stat. 1724), the National Institute of Standards and Technology shall be responsible for developing and proposing standards and guidelines needed to assure the cost-effective security and privacy of sensitive information in Federal computer systems.

(c) STUDY OF IMPACT OF FEDERAL PROCUREMENT REGULATIONS.—(1) The Secretary of Commerce shall conduct a study to—

(A) evaluate the impact of Federal procurement regulations that require that contractors providing software to the Federal Government share the rights to proprietary software development tools that the contractors use to develop the software; and

(B) determine whether such regulations discourage development of improved software development tools and techniques.

Reports. (2) The Secretary of Commerce shall, within one year after the date of enactment of this Act, report to the Congress regarding the results of the study conducted under paragraph (1).

(d) AUTHORIZATION OF APPROPRIATIONS.—From sums otherwise authorized to be appropriated, there are authorized to be appropriated—

(1) to the National Institute of Standards and Technology for the purposes of the Program $3,000,000 for fiscal year 1992; $4,000,000 for fiscal year 1993; $5,000,000 for fiscal year 1994; $6,000,000 for fiscal year 1995; and $7,000,000 for fiscal year 1996; and

(2) to the National Oceanic and Atmospheric Administration for the purposes of the Program $2,500,000 for fiscal year 1992; $3,000,000 for fiscal year 1993; $3,500,000 for fiscal year 1994; $4,000,000 for fiscal year 1995; and $4,500,000 for fiscal year 1996.

15 USC 5525. **SEC. 205. ENVIRONMENTAL PROTECTION AGENCY ACTIVITIES.**

(a) GENERAL RESPONSIBILITIES.—As part of the Program described in title I, the Environmental Protection Agency shall conduct basic and applied research directed toward the advancement and dissemination of computational techniques and software tools which form the core of ecosystem, atmospheric chemistry, and atmospheric dynamics models.

(b) AUTHORIZATION OF APPROPRIATIONS.—From sums otherwise authorized to be appropriated, there are authorized to be appropriated to the Environmental Protection Agency for the purposes of the Program $5,000,000 for fiscal year 1992; $5,500,000 for fiscal year 1993; $6,000,000 for fiscal year 1994; $6,500,000 for fiscal year 1995; and $7,000,000 for fiscal year 1996.

15 USC 5526. **SEC. 206. ROLE OF THE DEPARTMENT OF EDUCATION.**

(a) GENERAL RESPONSIBILITIES.—As part of the Program described in title I, the Secretary of Education is authorized to conduct basic and applied research in computational research with an emphasis on the coordination of activities with libraries, school facilities, and education research groups with respect to the advancement and dissemination of computational science and the development, evaluation and application of software capabilities.

(b) AUTHORIZATION OF APPROPRIATIONS.—From sums otherwise authorized to be appropriated, there are authorized to be appropriated to the Department of Education for the purposes of this section $1,500,000 for fiscal year 1992; $1,700,000 for fiscal year 1993; $1,900,000 for fiscal year 1994; $2,100,000 for fiscal year 1995; and $2,300,000 for fiscal year 1996.

15 USC 5527. **SEC. 207. MISCELLANEOUS PROVISIONS.**

(a) NONAPPLICABILITY.—Except to the extent the appropriate Federal agency or department head determines, the provisions of this Act shall not apply to—

(1) programs or activities regarding computer systems that process classified information; or

(2) computer systems the function, operation, or use of which are those delineated in paragraphs (1) through (5) of section 2315(a) of title 10, United States Code.

(b) ACQUISITION OF PROTOTYPE AND EARLY PRODUCTION MODELS.—In accordance with Federal contracting law, Federal agencies and departments participating in the Program may acquire prototype or early production models of new high-performance computing systems and subsystems to stimulate hardware and software development. Items of computing equipment acquired under this subsection shall be considered research computers for purposes of applicable acquisition regulations.

SEC. 208. FOSTERING UNITED STATES COMPETITIVENESS IN HIGH-PERFORMANCE COMPUTING AND RELATED ACTIVITIES. 15 USC 5528.

(a) FINDINGS.—The Congress finds the following:

(1) High-performance computing and associated technologies are critical to the United States economy.

(2) While the United States has led the development of high-performance computing, United States industry is facing increasing global competition.

(3) Despite existing international agreements on fair competition and nondiscrimination in government procurements, there is increasing concern that such agreements are not being honored, that more aggressive enforcement of such agreements is needed, and that additional steps may be required to ensure fair global competition, particularly in high-technology fields such as high-performance computing and associated technologies.

(4) It is appropriate for Federal agencies and departments to use the funds authorized for the Program in a manner which most effectively fosters the maintenance and development of United States leadership in high-performance computers and associated technologies in and for the benefit of the United States.

(5) It is appropriate for Federal agencies and departments to use the funds authorized for the Program in a manner, consistent with the Trade Agreements Act of 1979 (19 U.S.C. 2501 et seq.), which most effectively fosters reciprocal competitive procurement treatment by foreign governments for United States high-performance computing and associated technology products and suppliers.

(b) ANNUAL REPORT.—

(1) REPORT.—The Director shall submit an annual report to Congress that identifies—

(A) any grant, contract, cooperative agreement, or cooperative research and development agreement (as defined under section 12(d)(1) of the Stevenson-Wydler Technology Innovation Act of 1980 (15 U.S.C. 3710a(d)(1)) made or entered into by any Federal agency or department for research and development under the Program with—

(i) any company other than a company that is either incorporated or located in the United States, and that has majority ownership by individuals who are citizens of the United States; or

(ii) any educational institution or nonprofit institution located outside the United States; and

(B) any procurement exceeding $1,000,000 by any Federal agency or department under the Program for—

(i) unmanufactured articles, materials, or supplies mined or produced outside the United States; or

(ii) manufactured articles, materials, or supplies other than those manufactured in the United States substantially all from articles, materials, or supplies mined, produced, or manufactured in the United States,

under the meaning of title III of the Act of March 3, 1933 (41 U.S.C. 10a-10d; popularly known as the Buy American Act) as amended by the Buy American Act of 1988.

(2) CONSOLIDATION OF REPORTS.—The report required by this subsection may be included with the report required by section 101(a)(3)(A).

(c) REVIEW OF SUPERCOMPUTER AGREEMENT.—

(1) REPORT.—The Under Secretary for Technology Administration of the Department of Commerce (in this subsection referred to as the "Under Secretary") shall conduct a comprehensive study of the revised "Procedures to Introduce Supercomputers" and the accompanying exchange of letters between the United States and Japan dated June 15, 1990 (commonly referred to as the "Supercomputer Agreement") to determine whether the goals and objectives of such Agreement have been met and to analyze the effects of such Agreement on United States and Japanese supercomputer manufacturers. Within 180 days after the date of enactment of this Act, the Under Secretary shall submit a report to Congress containing the results of such study.

(2) CONSULTATION.—In conducting the comprehensive study under this subsection, the Under Secretary shall consult with approprite Federal agencies and departments and with United States manufacturers of supercomputers and other appropriate private sector entities.

(d) APPLICATION OF BUY AMERICAN ACT.—This Act does not affect the applicability of title III of the Act of March 3, 1933 (41 U.S.C. 10a-10d; popularly known as the Buy American Act), as amended by the Buy American Act of 1988, to procurements by Federal agencies and departments undertaken as a part of the Program.

Approved December 9, 1991.

LEGISLATIVE HISTORY—S. 272 (H.R. 656):

HOUSE REPORTS: No. 102-66, Pt. 1 (Comm. on Science, Space, and Technology) and Pt. 2 (Comm. on Education and Labor), both accompanying H.R. 656.
SENATE REPORTS: No. 102-57 (Comm. on Commerce, Science, and Transportation).
CONGRESSIONAL RECORD, Vol. 137 (1991):
July 11, H.R. 656 considered and passed House.
Sept. 11, S. 272 considered and passed Senate. H.R. 656 considered and passed Senate, amended.
Nov. 20, S. 272 considered and passed House, amended.
Nov. 22, Senate concurred in House amendments.
WEEKLY COMPILATION OF PRESIDENTIAL DOCUMENTS, Vol. 27 (1991):
Dec. 9, Presidential remarks.

O

A Special Issue of Electronic Networking: Research, Applications, and Policy

Accessing Information on the Internet

George H. Brett
Guest Editor

This special issue of Electronic Networking: Research, Applications and Policy focuses on the issues of networked information retrieval and on specific tools for successfully retrieving networked information. These topics are discussed widely on electronic lists on a daily basis. Networked Information Retrieval (NIR) is widely seen in the literature of information and library science as well as computing journals. NIR issues and tools are crucial to the future of information technologies. These are all different venues with various voices. This special issue is designed to provide a coherent and comprehensive discussion of these topics, a moderated forum if you will. But even as this special issue is completed, the topics, policy issues, and research questions regarding NIR are growing and expanding rapidly.

Indeed, the situation is much like the roadside of the southern United States which has been taken over by a rapid growing vine. The vine, kudzu, was brought to the United States many years ago to serve as a ground cover and as fodder for cattle. For the most part, the vine is spreading widely in rogue manner with little control. Kudzu is considered a weed by many. Yet, in Japan, where the root is used for cooking, kudzu is a valuable cash crop. In determining the value of this natural network, the abilities to perceive and to process resources and to manage growth are critical.

We face a similar challenge to our ability to perceive and manage resources in the current kudzu-like growth of the Internet. The Internet is rapidly becoming an interactive, real-time environment. Computing is done on the network in distributed, client-server modes with resources that are distributed worldwide. There are now more than 760,000 host computers worldwide that are connected to and using the Internet (Quarterman, 1992). Traffic from news groups alone generates more than 27Mbs of data per day (News.lists, 1992). The number of public resources that one can reach from any point on the network has grown as well. For example, currently there are more than 300 library catalogs or related services on the Internet. What we have is a tangled vine of networks and information resources. The Internet will become only more complex.

The people who use the network are another part of the growth problem. We see increasing numbers of people using the network who are not technologically informed. These individuals represent a broad spectrum found in the academy, from the humanities to pure sciences. If they use the network, it is to do specific functions such as electronic mail (e-mail). Often these persons lack technical sophistication and therefore require more support. Their demands will continue and increase. How can the network continue to provide them opportunities to do their research, teaching, publication, and other daily work with the resources of the Internet, at a level of informed independence?

We and they need appropriate tools. These tools must be integrated seamlessly into the networked environment, and just as seamlessly into our workplaces, our very desktops. The person who uses the network must be able to navigate the network to locate, access and use or manipulate information resources of every type.

Networked Information Retrieval is important for yet another reason. NIR has become the meeting grounds for two major cultures of information technologies: the information sciences (libraries) and the computing/networking technologies (computer centers). Events such as the formation of the Coalition for Networked Information (CNI) point to the positive aspects of these players' working together to create order from potential chaos. But such cooperative or combined ventures have been rare. In this

***George H. Brett** <ghb@concert.net> is Manager for Networked Information at the Center for Communications at MCNC in the Research Triangle Park, NC and concurrently also Consultant for Libraries and Networked Information to The University of North Carolina General Administration. Brett has been working since 1989 on issues of networked information resources with the EDCUOM Educational Uses of Information Technologies (EUIT) and with the Coalition for Networked Information (CNI). He also was a Citizen Delegate from North Carolina to the Second White House Conference on Libraries and Information Services. Brett is working toward his doctoral degree at the School of Information and Library Science at the University of North Carolina at Chapel Hill.*

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 10–12.

issue, we have provided writers a platform to represent their viewpoints. Differences are evident, but such differences should provide the launching pad for discussion.

In the opening article, Lynch and Preston explore issues of how we go about describing and categorizing resources found on the networks. They present a broad picture of problems that have already been encountered by both the library community (cataloging) and the network community (scale). They set a stage for us to evaluate the potential of future alternatives.

Hill and Neuman follow with recommendations for organizing directories of information resources. Hill's presentation of the X.500 directory services explains some of the rationale and problems likely to be associated with developing such directories. Neuman's Prospero is an approach in an environment that works within the personal workstation attached to the Internet. Such approaches offer one technique to help manage the growing resources available on the Internet.

Locating resources on the Internet can be a frustrating exercise. Deutsch and Scott present two solutions. Scott documents the HYTELNET system and how he has personally collected information about networked resources and he indexed them, and presents the information in a useful fashion with hypertext on a microcomputer. Deutsch, on the other hand, uses larger scale computing power. His automated system, Archie, actively goes out onto the network for information it uses to build a database. The database is then available for searching by the Internet community.

One must be able to access and use resources once they have been located. Kahle and Berners-Lee present their approaches for doing so. Berners-Lee's approach indicates incipient interoperability and offers a most interesting approach to access and manage Internet resources. In April 1991, Wide Area Information Server (WAIS) software was released to the public domain by Thinking Machines, Corporation. Since then, Kahle has actively promoted the development of clients and servers as well as the integration of WAIS into other NIR applications. Kahle describes how WAIS can be used successfully in a corporate environment. Berners-Lee writes how he uses the whole of the Internet as a World-Wide Web (WWWeb) of resources. His article indicates how interconnected NIR tools can be as he demonstrates how his application can utilize specific attributes of other NIR applications.

The final item in this special issue is an information sheet describing the Gopher Service, recently developed at the University of Minnesota. Gopher is especially useful as a browsing tool and it works cooperatively with, for example, Archie, WAIS, and the World-Wide Web.

The interoperability of networked information tools cannot be emphasized enough. There is little evidence that there will be only one brand of computer that will use a single operating system with one set of applications. In a world where computers were not connected together via networks, there was less importance attached to the exchange of data. But, as we create more connections, the ability to share data, information and programs becomes crucial. If interoperability is to take place, then well-defined standards and good implementation of these standards are critical.

The articles in this issue present various approaches that describe different points of view, but it should be noted that the applications described here already require interoperability. The way they work together is not so much on the level of the programs themselves as much on the exchanges of the information that they create. Both WAIS and WWWeb can use the data created by the Archie server or by Hytelnet. Recent electronic communications on the Internet indicate that these authors are working to further improve this exchange of data (Berners-Lee, 1992).

Another issue that plagues novice network navigators is how to find out what is on the network. For the moment, human intermediaries, such as reference librarians or colleagues, continue to help us locate those elusive resources. Lynch and Preston point out that we do not always need access to the entire universe of information. What we really require is a subset that applies to our particular interests or those of the community in which we work. Scott and Deutsch offer solutions that collect information for the user. Kahle and Berners-Lee then provide us with the tools to explore this collection. What we see are the beginnings of a more individual, more rational approach to identifying and accessing networked information.

We are just beginning to grasp the enormity of the impact of Networked Information Retrieval. For some time now the popular press has described the "information age." The 20% to 30% monthly growth of the Internet illustrates a demand for information resources on a global basis. But that demand is not for information resources alone. Rather, the increasingly more vociferous demand is for development of

adequate tools to identify, access, and use the resources—and, more to the point of this issue—for approaches that stress interoperability. Without the cooperative research, development, and application that interoperability implies, networks are no more than electronic superhighways tangled with rampant rogue kudzu.

References

Berners-Lee, Tim, Groff, Jean-François, & Cailliau, Robert. (1992). *Universal document identifiers on the network.* Geneva, Switzerland: CERN. UDI=file://info.cern.ch/pub/www/doc/udi1.ps

News.lists. (1992, February 12, 16:50:56). Total traffic through uunet for the last 2 weeks. UUNET Communications Services.

Quarterman, John. (1992, February 16). Internet growth. Com-priv (com-priv@psi.com) [list-serv]. February 16.

Describing and Classifying Networked Information Resources

Clifford A. Lynch and Cecilia M. Preston

The need for effective directories of networked information resources becomes more critical as these resources—online library catalogs, file archives, online journal article repositories, and information servers—proliferate, and as demand grows for intelligent tools to navigate and use such information resources. The existing approaches are based primarily on print-oriented directories, but print-oriented directories will not scale to support the future services that will help network users navigate tens of thousands of resources. The paper first explores the "user" perspective in various usage scenarios for employing a database of descriptive information to navigate or access networked information resources. It then considers specific data elements that will be required in a description of these networked information resources. Classification of networked information resources will ultimately rely on large-scale prototypes, coupled with a new generation of advanced information-seeking tools, and within the reality of economics.

The need for effective directories of networked information resources becomes more critical as these resources— online library catalogs, file archives, online journal article repositories, and information servers—proliferate, and as demand grows for intelligent tools to navigate and use such information resources. The existing approaches are based primarily on print-oriented directories such as the "Internet Accessible Guide to Library Catalogs" (St.George & Larsen, 1992) and the NSF-sponsored "Internet Resource Guide" (Partridge & Roubicek, 1989). Certainly, the network is used as a distribution medium for these guides, but, ultimately, most users of these tools make printed copies after transferring them to a convenient local machine. In the case of the "Internet Resource Guide," the file format is PostScript and the guide design is that of an updatable three-ring binder; so users can only browse or print images of print pages on bit-mapped displays.

This type of directory will not scale to support the future services that will help network users navigate tens of thousands of resources. While such guides represent an heroic attempt to fill short-term needs on a grass-roots basis (often with resources begged, borrowed, or stolen from other projects, or contributed by a sponsoring organization as a service to the networking community as a whole), they cannot solve the long-term problem.

Currently, much attention has been focused on access methods for directory information. Various factions argue that X.500 directories, specialized TELNET-access databases, Z39.50, a Wide Area Information Server (WAIS) directory of servers, or

***Clifford A. Lynch** <calur@uccmvsa [Bitnet]> is the Director of the Division of Library Automation at the University of California, Office of the President, Oakland, CA, 94612-3550 where he is responsible for the MELVYL ® information system, one of the largest public access information retrieval systems in existence, as well as the computer internetwork linking the nine UC campuses. He has been at the University of California in various positions since 1979. Dr. Lynch has also been involved in a wide variety of research and development efforts in the application of advanced technologies to information management and delivery, including work with computer networking, information servers, database management systems, and imaging technologies. Dr. Lynch received his Ph.D. in Computer Science from the University of California at Berkeley. He participates in several standards activities (including the NISO Standards Development Committee), is Principal Investigator of various research grants, and is the author of several books and over 50 published papers.*

***Cecilia M. Preston** <cpreston@info.berkeley.edu> is an independent information broker and a doctoral candidate at the University of California, Berkeley, School of Library and Information Studies, Berkeley, CA, 94720. She completed her BA from Glassboro State, NJ and her MLS from Rutgers University, New Brunswick, NJ, and has worked in both academic and corporate libraries.*

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 13–23.

some other new class of mechanism is the best way to provide directory services for networked information resources. We contend that each of these access methods will likely find a place as a means of locating networked information and that, in fact, there is no single right way to organize such a directory. Directory access mechanisms typically make sense only in the context of broader information access systems. In fact, we are already seeing the first generation of information access interfaces that employ networked information resource directory data as an integral part of a broader access system. Hytelnet (Scott, 1992), WAIS (Kahle & Medlar, 1991), and the Internet Gopher system (McCahill, 1991) are examples of such efforts.

The creation of records describing a network information resource, and making these records available to anyone who wants them, is not the same as paying for listing in a directory.

A great deal of glib discussion within the networking community has focused on the need to develop a network "yellow pages" to parallel the "white pages" (a directory of people and organizations) projects that are currently under development within the X.500 implementor community (Lang & Wright, 1992). This is often viewed as the goal for which the competing access methods are being championed. While superficially appealing, there are several problems with this concept. First, although it seems fairly clear how to describe a person or group, it is not clear how to describe the general instance of a networked information resource.

There is a major difference between white pages and yellow pages: In general, one searches the white pages for a known item (a named individual or company) to confirm or obtain an address or phone number; in the yellow pages, one searches for an unknown item. In fact, the printed yellow pages offered by the telephone companies offer poor guidance: Telephone company yellow pages are advertising and provide meager assistance in differentiating one service provider from another. Claims made in the yellow pages are made by the service provider; other tools, such as consumer reports, provide (not necessarily unbiased) evaluative information.

For example, the yellow pages provides little real assistance in locating a locksmith in a large city, and the user of such a directory will normally pick from one of a few large-display advertisements for lack of any better method of distinguishing one locksmith from another. Furthermore, telephone yellow pages need only provide minimal instruction on how to use the resources described in them—one dials the phone. And they are used by people, not machines, which means that a great deal of imprecision in classifying resources and in explaining how to use them (how and where the phone number is listed) can be tolerated; the reader will compensate for the variations in format. Users—and especially future networked information access systems, operating on behalf of people—will need a lot more than a simple electronic analog to the print yellow pages.

We believe that the real difficulty is devising the classification schemes for networked information resources and the specification of the data elements that users (both computer programs and people) of these resource descriptions will require. Once these problems are solved, it is relatively simple to present descriptions of networked information resources through a variety of access methods ("directories") such as Z39.50 and X.500. To be sure, conventions will be needed for creating and maintaining the databases that are presented through this diversity of access mechanisms; and transfer format standards for records containing descriptive data elements will have to be established.

Three types of information describe a networked resource. The first is factual: its name, who operates it, how to connect to it, and so on. The second is advertising: in this context, assertions about the resource made by its owner or operator that are not necessarily objectively derivable from the contents and services offered by the resource. The third class is evaluative: subjective information about the resource provided by third parties. The boundaries are murky; for example, "subject headings" are somewhat subjective.

This article focuses primarily on actual descriptive information for networked resources, although we will be somewhat liberal and will consider some forms of "subject access."

There have been several proposals for data elements to describe networked information resources. The primary characterization has its roots in descriptive bibliographic cataloging, and is summarized in Library of Congress (LC) Discussion Paper 49 (Library of Congress, 1991a) and its follow-on Discussion Paper 54 (Library of Congress, 1991b).[1] Although these papers provide a good beginning, many fields (for example, the Network Access Instructions) are not structured precisely enough to permit easy parsing by computer programs that might need to execute these access instructions on behalf of a user. In addition, the LC discussion paper provides for descriptive index terms and even a (not fully defined) element called "collection strength," but they do not provide for computer-processable content definition or evaluative information.

Agreement will also be necessary on whether the providers themselves or third-party directory compilers will create and maintain the various components of descriptions of information resources. Both models present problems. It is not clear that all, or even most, organizations that supply network information resources have the expertise to prepare appropriate descriptive records in the appropriate standard interchange formats. When universities discuss this problem, there is often a tacit assumption that university libraries can be relied on to provide the "cataloging" for networked resources. But, in general, this is simply untrue. Consider scientific data archives, or state data archives, or file transfer archives. This diversity requires third-party record creation services that can provide information records (presumably to be given away) to provider organizations which cannot build them using their own internal resources.

It is probably counterproductive for users to have to purchase descriptive (or factual) directory records [2], or to have license restrictions limiting their free flow in the network, or to have to access many overlapping and competing proprietary directory databases for factual descriptions. But the alternative is for the suppliers of information resources to fund the creation of factual descriptive records or for the overall user community to fund development of such descriptions as a community benefit.

The creation of records describing a network information resource, and making these records available to anyone who wants them, is not the same as paying for listing in a directory. Inclusion in directories raises issues of governance. (Do you pay to be listed? Does the manager of the directory feel that the nature of your resource—for example, "adult material" — is suitable for inclusion?) Separating the creation of descriptive records for networked resources from the inclusion of these descriptions in any specific directory (i.e., a database of such records, perhaps supplemented with advertising or third-party evaluations) avoids the dilemma of governance of "the" directory, and instead, allows a marketplace in directory entries — with equal access by all information resource providers, at least at this stage—to evolve.[3]

Some network information resources can be, to an extent, self-describing in an integral fashion, in that the same protocol used to access the resource can also be used to extract a description of the resource. For example, the EXPLAIN facility currently under development for Z39.50 contains a great deal of information that might be extracted into a resource directory entry for a given information server (Lynch, 1991). The information includes such server attributes as frequency of update, cost, number of records in the server, and identification of who maintains the server.

As information arrives from multiple sources, it is consolidated, ranked, and filtered, and periodically presented to the end-user.

In addition, the EXPLAIN facility provides a server with the capability to profile itself with respect to a classification scheme (discussed later). But it is important to recognize that a server cannot rate itself with respect to other servers (except by storing externally provided rating/comparison information, which logically is not part of the server's local self-description), nor can it objectively evaluate the quality and bias of its contents. Only an external directory database or a client performing computations on descriptive records (perhaps from multiple sources) has this ability. Evaluation and ranking are problematic because, to a great extent, they are value judgments. However, there are some useful forms of rankings based only on comparative statistics among servers.[4]

Some information resources—in particular, electronic journals, newsfeeds, sensor feeds, and mailing lists—do not have a well-defined method of storing a self-description that is accessible in the same way the resource is accessible. If a means for associating a description with such a resource is defined, it will, in essence, be part of a directory entry, though perhaps stored on a distributed basis. And unlike the EXPLAIN facility for information servers, it will probably be made available separate from the described resource, and through a separate access mechanism. We need to define where to store descriptions for resources that cannot naturally describe themselves.

This article explores the "user" perspective by examining various usage scenarios for employing a database of descriptive information to navigate or access networked information resources. These usage scenarios provide a basis for the development of requirements definition for describing and classifying such resources. We then consider specific data elements that will be required in a description of these networked information resources to facilitate the projected usage. One working assumption is that massive amounts of human effort by specialists will not be available to catalog these resources on an ongoing basis. Thus, we focus particularly on classes of resources for which computation can be used to develop much of the necessary resource description, building on, in some cases, human labor that has already been invested in creating the contents of the network resource. We conclude by examining classes of resources or user needs to identify resources where the computational methods that we propose do not seem to work well and where new insights still seem to be needed.

Usage Scenarios for a Directory of Networked Information Resources

Today, most use of directories of networked information resources follows one of two simple patterns. The most common is the known item search: One wants some information from server X (the "name" of X being known, at least to some degree of precision: "The Genebank server at the University of Houston," the "MELVYL®[5] system," "The University of California Online Catalog," "MELVYL.UCOP.EDU"), and it is necessary to determine the address[6] and some access information about that server. For example, it "speaks" Z39.50 at TCP port 210, or one sends a mail message in some specific fixed format, or one logs on via TELNET with a user ID "guest" and password "anonymous," possibly issuing a series of arbitrary and obscure commands to set terminal type and navigate various layers of front-end hardware and software before actually connecting to the resource desired.

Presently, the primary source for resolving such queries is either searching a printed directory or doing full-text searching on an ASCII version of such a printed directory. In fact, existing technology (and specifically, the existing unstructured textual description of resources) is largely adequate for known item searching. The two shortfalls are that descriptions do not contain all the details one might want, and that the lack of structure in the descriptions makes it hard to imbed use of these directory entries in larger access systems without substantial manual intervention. Both of these problems are abating as existing directory projects mature.

WUGATE at Washington University, the MELVYL system at the University of California, and the Colorado Alliance of Research Libraries (CARL) are systems that offer menus of remote networked information resources and then handle the minutiae of logon for their users. These systems were unable to use the directories available directly. Entries from the directories were hand transcribed. More recently, we believe second-generation tools like Hytelnet are doing at least semi-automated reformatting of directories entries from electronic copies of the existing print-oriented journals.

Also common (but today less common than known item searches) is a very limited form of subject search in which one requests resources such as "FTP archives containing the MIT X-Windows distribution." The result is a list of FTP servers from which the user selects one. (Today, the choice is often based on geographic proximity, which is not necessarily a good criterion in a networked environment; but the user has little else to go on in most cases.) Tools like the Archie service (Deutsch & Emtage, 1992) are used to support these types of searches. The user must know *a priori* that he/she is looking for an FTP site (i.e., the type of resource as well as the subject). It is only very recently that systems like the Internet Gopher are beginning to provide access to multiple types of resources (e.g., FTP archives, WAIS servers, TELNET-accessible systems).

Imagine that we are a few years in the future: A researcher is interested in tracking information on treatments for arthritis and enters a request into his or her workstation to be kept abreast of developments in this area. The software on the workstation determines the general area of discourse (health and

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

Business Reply Mail

First Class Mail Permit No. 66 Westport, CT

POSTAGE WILL BE PAID BY ADDRESSEE

Meckler
11 Ferry Lane West
Westport, CT 06880-9760

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

Business Reply Mail

First Class Mail Permit No. 66 Westport, CT

POSTAGE WILL BE PAID BY ADDRESSEE

Meckler
11 Ferry Lane West
Westport, CT 06880-9760

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

Business Reply Mail

First Class Mail Permit No. 66 Westport, CT

POSTAGE WILL BE PAID BY ADDRESSEE

Meckler
11 Ferry Lane West
Westport, CT 06880-9760

For more information

Call 1-800-635-5537 in the U.S. or Canada or write to Meckler, 11 Ferry Lane West, Westport, CT 06880, U.S.A. Overseas call 203-226-6967.

biomedical topics, in this case) and then invokes an entry vocabulary (perhaps the National Library of Medicine's UMLS (Lindberg & Humphreys, 1990), or some descendant of it) to normalize the terms in the query supplied by the user. There may be some dialogue with the user to clarify or refine the request, or to determine the user's budget, how comprehensive the results should be, and perhaps other qualifications for the type of information requested. (Is the user a physician here? Or does the user want popular or research-level articles? Does the user want articles in languages other than English?)

At this point, the workstation must determine the relevant networked information resources, how frequently they are updated, the searching mechanism for each resource, and any cost for searching each of the resources (consulting local and remote databases describing and evaluating available resources as necessary). Having determined the relevant resources, the workstation will then execute a set of heuristics (perhaps incorporating further interaction with the information-seeker) to decide which resources to search and how often, and proceed to acquire information on behalf of the user. As information arrives from multiple sources, it is consolidated, ranked, and filtered, and periodically presented to the end-user.

There are complex relationships between a class number profile of a library's collection and the statements contained in a conspectus entry.

There are several important details about this scenario. First, the user either does not see directory entries at all or sees them only in a highly filtered form. For example, the workstation might ask the user for input about which of two or three highly ranked resources are most important. The system might ask the user to confirm use of a particularly costly information resource. Or the system might select a resource that appears relevant but with which the user is unfamiliar (not having used it before, as the system knows from inspecting history files), and the user may ask who manages it. The user may wish to know about biases of the information provider.

Databases of public information on the environment offered by a major oil company or by the Sierra Club might show different perspectives. Here, a computer program—not a human being—is the primary consumer of the descriptions of information resouces. This program decides what information, if any, about a resource with which the end-user will be troubled, based on inspection of the resource description, in the context of the user's known biases, history of information access, budget, and other parameters.[7]

Second, the workstation employs some universal taxonomy (at least within a certain universe of discourse) to select the relevant information resources. Such a taxonomy is the only means of classifying information resources in such a way that programs can make relative evaluation of one resource versus another. This taxonomy is almost certainly related to the entry vocabulary applied to the user's initial request to match the vocabulary used by the information-seeker to the vocabularies used to describe the items stored in the various available information resources.

The objective, then, must be to develop data elements that can be part of a description of a networked information resource which can support the resource identification and evaluation requirements of future intelligent software agents that seek out and organize information on behalf of a user.

We distinguish here two other types of problems in locating information resources, sometimes called resource discovery, which are related to, but different from, the focus of this article. One determines an appropriate instance of a class of networked resource—for example, the location of the nearest free printer, the nearest authentication server, the nearest name server, or the nearest directory server for networked information. This type of problem adds some complexities that are beyond the scope of this article, particularly if a good choice is to be made. But this can be viewed as a closely related problem to known item searching, in that the set of categories of networked resources is relatively small and well defined, and membership of a resource in a class is unambiguous.

The other form of resource discovery, which has been pioneered by the work of Mike Schwartz at the University of Colorado (Schwartz, 1991), deals with the situation when descriptions of available resources do not exist, and it is necessary to resort to

heuristics simply to determine a set of possible resources to be examined further, most probably by the end-user. Schwartz's scenario merely suggests "likely" possibilities. This is an interesting problem but has a very different focus than our work. However, some of the techniques Schwartz proposes may be useful in developing subjective, evaluative information about networked resources that can be used as an adjunct to the classification information discussed here. These techniques seem well suited to acquiring information about relative interest in various resources by members of modest-sized communities of network users with which the information-seeker may declare an affinity of interests.

Classification Schemes for Networked Information Resources

Consider the case of an online catalog [to start, assume specifically a monographic database, not the abstracting and indexing (A&I) databases currently used to extend many online catalogs]. This database represents a library collection. There are currently several widely understood and accepted means of profiling a library collection and, hence, profiling an online catalog as a surrogate for the collection. These include:

1. Library of Congress (LC) class number profiling. This method has been used for some time for collection characterization (Paskoff & Perrault, 1990; Branin, Farrell, & Tiblin, 1985; and Evans, Gifford, & Franz, 1977). It is possible to assign both absolute and relative counts of the number of matching objects. One deficiency here is that objects typically have a single call number (and thus class number). Thus, an object is only profiled once, though it may be relevant in several categories. Another problem is that class number assignment decisions are sometimes made relative to overall scope of a specific collection and not in the abstract.
2. LC subject headings (LCSH). It is possible to count (both in absolute and relative terms) the number of items in the collection which have a specific LC subject heading. It is also possible to relate these items to LC class numbers through statistical correlations. This procedure can be used to associate multiple-class numbers with an object that has multiple subject headings. These assignments can also be weighted in various ways (e.g., the class number from a call number or from primary subject heading will count more than those derived from other subject headings).

   In addition, a hierarchy may be imposed through use of the broader term/narrower term relations that are incorporated in the LC subject headings. The hierarchy can be used to help cluster headings or derived class numbers, or to avoid the set of class numbers assigned to a work by adding "weak" class numbers corresponding to more or less specific subject headings than those assigned with the number of broader or narrower term relationships separating the derived from the assigned terms determining the degree of weakness.
3. Conspectus. The conspectus (Gwinn, 1985; Gwin & Mosher, 1983; and Ferguson, Grant, & Rutstein, 1987) is a means of characterizing the belief and intentions of a library regarding its collection coverage. It is a form that is completed by the collection development staff providing their evaluation of both the strengths of the collection and the collection policies of the library in each subject area. Obviously, these completed forms can form a database. A library can use a conspectus entry to declare an intent to collect comprehensively in a specific area, or to state that a collection is felt to be comprehensive in a specific area. The subject breakdown in the conspectus can be mapped to class numbers, and thus to subject headings.

There are complex relationships between a class number profile of a library's collection and the statements contained in a conspectus entry. Statistical analysis of the class number profile of a collection (including a consideration of acquisition dates) has been used to validate conspectus entries with a substantial degree of agreement (McGrath & Nuzzo, 1991; Mosher, 1985). While a characterization of the collection by class number (or subject heading) analysis is perhaps the best way to characterize the past and the present, and even to determine trends by looking at the profile relative to publication date (or, more accurately, acquisition date), the conspectus is the only way for a library to state its intentions.

By using LCSH as an entry vocabulary, it should be possible to determine which online

catalogs (and thus library collections) are strong in a given subject area. The entry vocabulary aspect of LCSH can be used to compute authoritative headings from user-supplied terms and then, statistically, class numbers if these are desired. There are two ways to look at this information: by absolute collection count or by percentage of the collection (which might be viewed as an indicator of collection emphasis, at least in some libraries). Conspectus data can also be used as an additional indicator of collection focus, and this can be correlated with the LCSH terms. To compare libraries, one can look at both conspectus data and the number of holdings for the class numbers or subject headings of interest.

Libraries are, in many cases, already investing in conspectus statements. The remaining analysis described above is primarily a matter of computer time to characterize the collection statistically and develop correlations. There is a good deal of research required that will determine how often the collection statistical profile needs to be recomputed as the collection grows and changes. There are also a number of messy technical issues, such as the assignment of LC call numbers (and hence class numbers) that are different from those the Library of Congress assigns to a work (quite common in special libraries). The methodology of characterization needs both research and experimental validation. But the basic capability for profiling already exists.

Similar collection profiling can be done based on any well-structured subject classification scheme, such as MeSH (Medical Subject Headings) or the INSPEC thesaurus, or the Defense Technical Information Center (DTIC) or National Aeronautics and Space Administration (NASA) classification schemes. Furthermore, if a sufficiently large collection of records exists somewhere that is classified under any two schemes, statistical analysis should be possible, with minimal human review, that will provide a reasonably accurate mapping table from one scheme to another. Such mapping tables could be used to bridge the gap between a query that is processed through one entry vocabulary and a resource that is profiled using a different classification scheme. Of course, it is also possible to develop these mapping tables through human intellectual analysis.

It is important to understand the approximate and statistical nature of this classification approach: It will work best, we believe, with fairly broad groupings such as Library of Congress class numbers. It is not necessarily a panacea that will resolve the problem of different classification vocabularies at the level at which the user is actually searching the specific information resources in question. But it should be useful in guiding choices among information resources.

A final point should be made about multiple classification schemes. It is almost certainly best to store the profile of a collection relative to the classification scheme that was used to catalog it. If a user of the directory wants to map it to another classification scheme (which will invariably lose some precision), a generally available mapping table can be used. Such mapping tables will likely improve over time, and this way no information will be permanently lost. As an amenity, it would be straightforward to support a primary profile using the collection's indigenous classification scheme, and secondary profiles in other classification schemes that are precomputed periodically (as correlation tables are improved) to reduce computational load on users of descriptive data.

It should be possible to extend this method of statistical collection profiling to serials collections by analyzing the LC class numbers and/or subject headings assigned to periodicals held in a library's collection. Another option, when the individual articles in journals are covered by an A&I file such as MEDLINE®, would be to perform more detailed analysis on the contents of each serial based on the articles that have appeared in it, and then map from the A&I files classification schemes as required. This scheme would not be perfect. For example, if the contents of a given journal shifts significantly from one year to the next, it will be very difficult to perform statistical characterization to reflect anything other than the "average" contents of the journal.

Analysis does have the advantage of accommodating a shifting journal focus, however, without relying on recataloging. Obviously, this type of analysis can be made on any journal in a library's collection (including electronic journals). Finally, a specific electronic journal is amenable to automatic indexing as a means of describing content, although time horizons need to be carefully considered.

It is probably most appropriate to "chunk" electronic journals or listservs or other network newsgroups chronologically (with the chunk size either corresponding to a specific period of time or a specific number of megabytes of traffic) and to perform automatic indexing at the chunk level. Research is needed on appropriate chunk size. An additional possibility that should be investigated is a variable chunk size, where chunk boundaries are determined by significant shifts in the terms that are identified by an automatic indexing process. It is

also important to consider the selection of the "universe" relative to which automatic indexing should occur—a single text stream (e.g., a newsgroup), the newsgroup considered in the context of a number of topically related newsgroups, all newsgroups,... The larger the "universe," the more complex the automatic indexing process will become.

Classification and Collection Granularity

One problem with this approach is that it depends on LCSH or some other classification scheme (perhaps in a subject-specific area) being universally understood. Essentially, each information resource characterizes itself (statistically or by analysis) relative to this classification scheme; clients then examine descriptions of various network information resources based on the nearness of a query to this classification profile.

Difficulties arise in two areas. The first occurs when the query or the collection is too specific to be characterized accurately relative to a classification scheme. Here there is a tension: One does not want too many classification schemes since this will "balkanize" the information resources on the network. Unless they are familiar with each highly specific classification scheme used, clients will be unable to assess information resources that have been characterized according to the various schemes. Of course, there is the possibility of developing a meta-classification scheme that provides a higher level framework for the more specific schemes: Imagine LCSH, enhanced with a series of more subject-specific schemes that further qualify certain "most specific" LCSH as narrower terms.

At least in theory, however, the development of specialized subject classification vocabularies that are in some sense "attached" to the LCSH (or other "major" classification systems) presents no problem. But some agency must take the responsiblity to develop and maintain the scheme in each area, and the schemes much be consistent and consistently linked to the more generally accepted schemes such as LCSH.

The second area of difficulty is more serious. If one considers the online catalog for a library collection, a great deal of intellectual effort has already been expended in cataloging the collection. The classification of this collection as an Internet resource essentially builds on this labor by performing statistical analysis on the MARC records that characterize the library's collection.

An FTP archive is in some sense the antithesis of a library collection. In a library collection, the items comprising the collection are considered of enough lasting value to justify the intellectual investment to catalog them in a structured fashion (and to assign Library of Congress subject headings). Our proposal for classifying online catalogs builds on this effort. In an FTP archive, the collection consists of contributed programs that have not been cataloged according to any uniform scheme. In fact, in many cases, the ephemeral nature and low unit value of the programs or other files comprising the archive suggest that the cost of cataloging is unsupportable. Here, it may be impossible to derive a "collection profile" by statistical analysis of cataloging records. At best, the manager of the FTP archive will be able to indicate the collection policy by filling in a conspectus-like statement (relative to a very specialized subject vocabulary).[8]

Another, related, problem can be illustrated by library collections: The statistics of subject heading (or class number) occurrence do not reflect collection strength in some important particulars. A library may hold a special collection (even one scanned into electronic form, and thus network-accessible) that is described only by a single collection-level record. It may be possible to resolve this problem by weighting collection-level records heavily in developing statistical profiles for library collections.

Another possibility is that a library might hold extensive individually catalogued items, the call numbers of which are not assigned within the class number schedule in the way that certain key authors (Shakespeare, James Joyce, U.S. presidents, etc.) are. Thus, this particular collection strength may be externally invisible when mapped relative to a classification scheme that is widely understood. The upshot of these problems is that it will be necessary to supplement structured statistical profiles with uncontrolled free text descriptions of special collections in descriptive records.

Other Issues in Nonbibliographic Information Resources

Currently, numerous data archives house extensive collections of material such as remote sensing imagery. These data archives can be treated much like online catalogs for the purpose of constructing resource descriptive records. Here, the individual datasets are analogous to books. The records are presumably housed in some type of local or disciplinary catalog. In fact, a central catalog might describe data

at multiple sites, such as the NASA NODIS system,[9] or the centralized catalog might even represent merely a "contact point" for accessing catalog records that are distributed along with the datasets themselves. The records are much like records in an online catalog, and, by examining classification information in the dataset descriptive records, one can characterize the archive.

But here, again, one encounters possible problems with the classification scheme: Can it differentiate by remote sensing platform or type of sensor, allowing a seeker of resources to find only data archives containing LANDSAT thematic mapping data? The severity of this problem may depend ultimately on the number of distinct data archive sites that come to exist on the network.

There is a related issue in classes of resources. A user interested in the author Lewis Carroll might be able to locate libraries with collections of books by and about Carroll, but might really be interested in locating data archives containing ASCII versions of Carroll's books. In some sense, the electronic text as an element of the data archive perhaps should have more "weight" or special status in that, to many network users, this text has a special, or at least different importance because it is immediately accessible. Perhaps the solution here is to profile resources to differentiate those that can supply network-accessible primary information.

In the case of the user seeking the full text of *Alice in Wonderland*, limiting the search to only data archives might at first consideration seem reasonable. But it seems clear that over the next few years, libraries will increasingly link full text or bit-mapped images deliverable over the network to the existing records in their online catalogs and A&I databases, thus taking on roles as data repositories.

More work may be needed to determine the best way to reflect this duality between primary material that is directly network-accessible and that which is only described in a network-accessible resource but has a separate physical existence. Hybrid cases, such as online catalogs of material an institution is willing to schedule for scanning on demand (and then deliver over the network—but where the primary data may take days, rather than seconds or minutes for a client to retrieve), present additional possible complications, as they suggest that network accessibility is really a continuum ranging from material stored online through "nearline" tape library robots and optical storage jukeboxes all the way to scan-on-demand facilities.

A final problem is the difficulty of classifying general reference databases such as the electronic versions of the annual *CIA World Factbook* or electronic encyclopedias. In a sense, this is an old problem familiar to those who have considered subject classification for reference works in libraries and the way this cataloging can mislead users. A user searching for information on the economy or history of a given country might miss some of these more general resources because they are not specific to the country in question, unless very large numbers of "subject headings" or indexing terms are applied to these electronic reference resources. For example, one might want to assign not only some broad terms, but many or all of the subordinate (e.g., narrower) terms in the classification scheme being used. The user may be particularly interested in locating such reference databases simply because they can provide immediate, network-accessible information to satisfy the user's query.

Conclusions

It seems clear that the problem of properly classifying networked information resources is difficult and far from entirely solved. Large-scale prototypes will be needed to provide insight into the open questions. Furthermore, these prototypes, to be properly understood, will have to be coupled with a new generation of advanced information-seeking tools, for they cannot be evaluated outside of that context. Hopefully, new efforts like the Coalition for Networked Information's (CNI) TopNode project (CNI, 1991), which seems to be focusing on creating databases rather than simply on access mechanisms, will help move our understanding forward.

The vision of a single, simple unified directory of networked information resources through a technology such as X.500 seems somewhat chimerical, if this directory is to support much beyond the simple known item lookup function. This article shows that even the structures for relatively objective resource descriptions are quite complex; they are large and require considerable computation and expertise to develop.

But beyond the complexity of objective statistical and extracted free-text characterization, there is the reality of economics. Subjective or mixed objective/subjective characterizations of networked information resources have fungible value. These evaluative descriptions will not be broadcast on the network for all to use freely.[10] Instead, it seems likely that a user's workstation, in identifying resources to

use in satisfying a query, will ask several resource identification servers for advice. In many cases, this advice may not be free. Responses to these queries will then be integrated against a local database of user history, biases, and preferences, and further pruned by heuristics that apply budgetary, timeliness, and comprehensiveness constraints. At this point, the user's workstation would begin to query the resources themselves, most likely incorporating adaptive feedback to further refine the set of resources to be queried on an ongoing basis.

This is a likely scenario for the real future of networked information resource directories. It will not be unified. It will not be simple. And it will develop in an evolutionary fashion as we better understand issues of classification.

Acknowledgments

We would like to thank Nancy Gusack for her editorial assistance; Michael Buckland (University of California, Berkeley), George Brett (MCNC Center for Communication, University of North Carolina), the CNI TopNode team: Ron Larsen (University of Maryland, DC), Art St. George (University of New Mexico), Craig Summerhill (CNI), Peggy Seiden (University of Pennsylvania), Brewster Kahle (Thinking Machines Corporation), and Mike Schwartz (University of Colorado); and the members of the University of California, Berkeley, School of Library and Information Studies Seminar on Networked Information, 1991–1992, for helping us to develop the ideas here, though we must, of course, take responsibility for all errors.

Notes

1. There is also a draft RFC giving an X.500 scheme for the RFC 49 data elements (Weider and Knopper, 1991).

2. This is not true for evaluative information. Network users might pay well for access to databases that provide high-quality evaluations of the quality and cost-effectiveness of other network resources. Such evaluative records will, in some cases, be valuable, proprietary assets.

3. This is not to say that some years hence we might not face a monopoly-type problem when some directory provider emerges that becomes dominant in the marketplace, and perhaps requires regulation to assure information providers equal access to listing in that provider's directory.

4. A basic observation about the "automatic" indexing that has been proposed for textual resource descriptions provided by resource operators is that such automatic indexing can only be done external to a given resource's self-description. The keyterms to be assigned to a given resource under typical automatic indexing IDF (inverse document frequency) weighting are those that appear frequently in the descriptions of only a few resources, and infrequently in the descriptions of other resources in the description set being indexed (Salton and McGill, 1983).

5. MELVYL is the registered trademark of The Regents of the University of California.

6. In the case where the exact name—for example, MELVYL.UCOP.EDU—is known, obtaining the IP address is by straightforward use of the Internet Domain Name System (DNS). The real problem is moving from a generic name or nickname to a precise DNS name after which the DNS can provide the IP address.

7. In fact, the processing that the client must do to select properly from potentially appropriate network information resources is enormously complex, and appropriate algorithms and heuristics deserve a great deal more attention than they have received to date. For example, some resources may be available only to closed communities (of which the user may or may not be a member), or may be available at discounted rates to members of such communities. Some resources may be at least temporarily unavailable due to network outages, or available only over slow or unreliable links. Diversity of sources (or diversity of viewpoints) may be an important selection consideration, as many multinational sources, or minimization of overlap.

8. Perhaps it may be possible to infer the contents of FTP archives statistically by analyzing certain files that are part of the archive (i.e., documentation files), as well as by doing keyword extraction from the comments in computer programs. But this will likely lead to a relatively sloppy and low-quality characterization of the archive's contents. In addition, descriptive records for networked information resources will not address the problem of locating something of interest with an FTP archive. It will only tell a client that a given FTP archive is liable to contain some information of interest for a given query.

9. To explore this system, one can connect to the MELVYL system and then issue the command "USE NASA."

10. In fact, even the objective characterizations may, in many cases, be provided by third parties.

OCLC, for example, has at its disposal the databases, computational resources, and expertise to derive statistical characterizations of the collections of the majority of library collections in the United States. Thus, by extension, it can compute descriptions of these libraries' online catalogs as network information resources. It seems likely that if OCLC offers such a service, they might well try to maintain control of the resulting database.

References

Branin, J. J., Farrell, D., & Tiblin, M. (1985). The national shelflist count project: Its history, limitations, and usefulness. *Library Resources and Technical Services, 29*(4), 333–342.

Coalition for Networked Information. (1991). *The TopNode for networked information resources, services, and tools.* Call for statement of interest and experience. (Available from Craig Summerhill, Systems Coordinator, CNI, 1527 New Hampshire Ave., NW, Washington, DC 20036)

Deutsch, Peter, & Emtage, Alan. (1992). The Archie system: An Internet electronic directory service. *ConneXions 6*(2), 2–9.

Evans, Glyn T., Gifford, Roger, & Franz, Donald R. (1977). *Collection development analysis using OCLC archival tapes: Final report.* ERIC Document ED 152 299. Albany, NY: Office of Library Services, State University of New York.

Ferguson, Anthony W., Grant, Joan, & Rutstein, Joel. (1987). Internal uses of the RLG conspectus. *Journal of Library Administration, 8*(2), 35–36.

Gwinn, Nancy E. (1985). Cooperative collection development—national trends and new tools. *Art Documentation, 4*(4), Winter, 143–147.

Gwinn, Nancy E., & Mosher, Paul H. (1983). Coordinating collection development: The RLG conspectus. *College and Research Libraries, 44*(2), March/April, 128–139.

Kahle, Brewster, & Medlar, A. (1991). An information system for corporate users: Wide area information servers. *Online Magazine, 15*(5), September/October, 56–60.

Lang, R., & Wright, R. (editors). (1992). *A catalog of available X.500 implementations.* RFC 1292. Palo Alto, CA: SRI International & Berkeley, CA: Lawrence Berkeley Laboratory.

Library of Congress. (1991a). Discussion paper #49. *Dictionary of data elements for online information resources.* Washington, DC: Library of Congress.

Library of Congress. (1991b). MARBI discussion paper #54. *Providing access to online information resources.* Washington, DC: Library of Congress.

Lindberg, Donald A.B., & Humphreys, Betsy L. (1990). The UMLS knowledge sources: Tools for building better user interfaces. *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care,* Washington, DC, November 4–7, 1990, Volume 14 (pp. 121–125). Los Alamitos, CA: IEEE Computer Society Press.

Lynch, Clifford A. (1991). *Towards an EXPLAIN facility for Z39.50.* Draft of 9/4/91, posted on Z3950IW@NERVM.BITNET on 9/05/91.

McCahill, Mark P. (1991). About gopher. Anonymous FTP to <boombox.micro.umn.edu> under / pub/gopher directory tree.

McGrath, William E., & Nuzzo, Nancy G. (1991). Existing collection strength and shelflist count correlations in RLG's conspectus for music. *College and Research Libraries,* 52(2), March, 194–203.

Mosher, Paul. (1985). The nature and uses of the RLG verification studies. *College and Research Libraries News, 46*(4), July/August, 336–341.

Partridge, Craig, & Roubicek, Karen. (1989). *Internet resource guide.* Cambridge, MA: NSF Network Service Center.

Paskoff, Beth M., & Perrault, Anna H. (1990). A tool for comparative collection analysis: Conducting a shelflist sample to construct a collection profile. *Library Resources and Technical Services, 34*(2), 199–215.

Salton, Gerard, & McGill, Michael J. (1983). *Introduction to modern information retrieval.* New York: McGraw-Hill.

Schwartz, Michael. (1991). Resource discovery and related research at the University of Colorado. *ConneXions, 5*(5).

Scott, Peter. (1992). Hytelnet as software for accessing the Internet: A personal perspective. *Electronic Networking: Research, Applications, and Policy,* 2 (1), 12–22.

St. George, Art, & Larsen, Ron. (1992). *Internet accessible library catalogs and databases.* (Available from STGEORGE@NMB.BITNET)

Weider, Chris, & Knopper, Mark (1991). *Schema for information resource description in X.500.* IETF OSI-DS Working Group. RFC Draft #49. Ann Arbor, MI: Merit Network.

The X.500 Directory Service: A Discussion of the Concerns Raised by the Existence of a Global Directory

Julia M. Hill

The X.500 Directory Service is one of the most important tools ever produced for network users. It is the enabling mechanism for a revolution in communications among people worldwide. Initiating the service, however, can be fraught with problems—not the technical challenges of creating a globally distributed service with locally managed controls, but concerns raised by the very existence of a worldwide database of information relating directly to individuals. Opportunities opened up by the use of the Directory are inevitably accompanied by the possibility of misuse. Individual subjects of the information have divided views. They earnestly wish for easier contact with colleagues and others worldwide, while entertaining in varying degrees a fear of invasion of privacy or a violation of personal rights. Managements taking responsibility for their staff and students are reacting with caution to requests for information for inclusion in the Directory. These concerns must be taken seriously, or the service will fail—either by not reaching the critical mass that will make it useful, or by quickly becoming out of date and therefore irrelevant. Prospective Directory Service managers must take considerable care to present the service in a reassuring way to their subjects and administrators, to convince them that the benefits greatly outweigh the risks, that controls exist, and that responsible Directory use will benefit the world network community.

The Case for Using X.500

The X.500 Directory Service provides a mechanism for finding information—information about people, organizations, services, network hardware, and more—in the global network environment. People now spend a great deal of time searching manually for names and addresses, specifically electronic mail names and addresses. They consult postmasters, computer center advisers, administrators, and directories. Yet if one queries the Directory from a workstation on the desk, information about people and other network resources can be instantly available. Each organization manages its own part of the Directory, storing information about as many of its staff and students, working groups, committees, and other useful resources as it wishes. It can be used to find people, job holders, or committee members. It can hold a map of the campus or photographs of the staff. The aim is that the Directory will become so easily accessible that people will use it without comment, in the way they now lift a telephone and make a call or ask for directory assistance. People will want to communicate with colleagues, or locate information in an efficient manner, without having to know how the Directory works. They should be able to consult it as easily as glancing at a list of local contacts pinned to the wall by the desk.

The Directory Service, commonly referred to as X.500, is an international standard, ratified by the International Organization for Standardization (IS) and the International Telegraph and Telephone Consultative Committee (CCITT) in 1988.[1] Based on a series of recommendations (CCITT X.500/ISO/ 9594), the Directory is an Open Systems Interconnection (OSI) Application layer standard and "defines one global directory that will be logically centralized (but physically distributed) across the numerous nodes that will interconnect to create the global network" (Planka, 1990, p. 95).

The service is intended to be global in the fullest sense of the word, encompassing not only academic institutions but also commercial companies,

Julia Hill *<j.m. hill @cluster.heriot-watt.ac.uk> graduated in pure mathematics from the University of Edinburgh in 1965. After some years in data processing, she joined Heriot-Watt University in 1979 as a systems programmer. Her present post is Network Services Manager for the University, with responsibility for planning and maintenance of all campus networks. She joined the UK X.500 academic pilot group at the outset, and leads negotiations with the UK Data Protection Office for the group. Mrs. J.M. Hill can be reached at Computing Services, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS Scotland. Telephone: +44314513267.*

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 24–29.

businesses of all kinds, manufacturers, and service providers. All these organizations will eventually be able to join the Directory. Some of the largest network users, including firms such as Hughes Aircraft Company, Rockwell International Corporation, TRW, and Xerox Corporation, are investigating or implementing X.500 services (Cope, 1991, p. 1). Network users are interested in the Directory because X.500 is the only worldwide standard for an electronic mail directory (Cope, 1991, p. 32).

Complementing the Directory, the Message Handling Service (X.400) defines the use of X.500-based services to provide "user-friendly naming, distribution lists, recipient capabilities, and authentication" (Planka, 1990, p. 94). The X.400 service can use the Directory automatically for searches, list expansion, and delivery instructions. Electronic mail is becoming an essential feature of everyday work life as more organizations provide workstations for increased numbers of their personnel. Electronic mail addresses, however, can be difficult to find if one is trying to make the first contact with another network user. The Directory potentially can save a great deal of time. Electronic mail addresses will be maintained automatically by the same software that registers their users for access to their organization's computer systems.

Probably the greatest concern expressed in an academic institution is management's reluctance to allow the development of any directory data.

The X.500 standard provides the structure for describing objects associated with a network. However, it does not specify what information about the objects will be made available to the Directory. Current projects in the United Kingdom, Europe, and the United States are exploring the technical and organizational issues of X.500 Directory Services. For example, the UK Academic X.500 pilot project provides an X.500 infrastructure to the university community (Dempsey, 1991, p. 49–50). The Cooperation for Open Systems Interconnection Networking in Europe (COSINE) effort is aimed at coordinating Directory pilot projects to develop a European Directory Service (Planka, 1990, p. 35). The New York State Education and Research Network (NYSERNET) has implemented a "white pages" directory for locating network users (Dalton, 1991, p. 35). The relative technical immaturity of the current Directory Service confronts network service providers and managers with many challenges. Yet organizations such as the Internet Activities Board's Internet Engineering Task Force are developing solutions to current technical shortcomings.

Another set of problems, however, is only beginning to be addressed. These problems range from getting an organization to commit itself (with appropriate resources and policies) to developing the Directory to facing issues related to personal privacy, security, and data integrity. Subsequent sections of this paper will detail some of the organizational, management, and privacy issues that a Directory Service engenders.

Introducing the Service to Management

The desire to participate in the Directory does not, of course, make all useful network information magically appear. It is no trivial matter to set up even a local directory information base that can be part of the larger distributed database forming the logical Directory. Appreciable amounts of staff resources will be needed to collect the information and initiate management procedures. Hopefully, the potential benefits will encourage organizations to set up their own part of the Directory, link it to the global service, and devise systems for integrating Directory management into everyday activities.

Maintaining the security of the stored information is absolutely vital if management and subjects of the information are to have confidence in the Directory Service. The 1988 X.500 standard defined authentication procedures (e.g., verifying the identity of a user by checking passwords) but did not provide access control guidelines specifying which user can perform what operations on what data. Today, however, access control is receiving attention: "Since both authentication and access control are considered essential components of a Directory, which certainly will contain restricted-access information, ISO and CCITT are currently defining a Directory Access Control Scheme" (Planka, 1990,

p. 100). Unless everyone is convinced that authorization, access control, and authentication, are being maintained, the Service will fail. If only a small proportion of an organization's members have Directory entries, the Service will not reach the critical mass necessary to swing the balance in its favor. If lack of commitment on the part of management leads to starvation of the resources needed to maintain the Directory information, such information will become useless.

The first step in introducing the Directory, by what could be called an internal salesperson for the Service, should be to approach the head of the organization. Support from people in the upper hierarchy will be of great value. Important people with whom to discuss the proposed service include the company secretary, public relations officer, personnel manager, and, if appropriate, the data protection officer. The organization's librarian, publications manager, telephone manager, and computing service director will also have a keen interest in the Directory. The necessity of gathering initial data and subsequent updates form primary sources, and the ultimate aim of creating the Directory as a master file, make the help of all these people vital. Some units of the organization may collect sensitive or confidential data, which should not be included in the Directory Service. Each responsible person has a master copy of his or her own files. Consequently, a major task in setting up the Directory is finding and merging all the appropriate sources and deciding which master takes precedence. This will prove to be one of the main problems in administering a full Directory Service.

Most organizations find the most obvious source of suitable information to be their internal phone book. Thus, the initial entries will most likely include name, telephone number, room number, department and electronic mail address. This will form an excellent basic service for any organization, and as confidence and realization of usefulness increase, more information can be added—often at the request of the subjects themselves.

A minimum entry for each individual in an organization is highly desirable for a Directory Service to function usefully. However, organizations may have some people with a real fear of being visible in the Directory. It will be necessary to provide an *ex-directory* option, either completely excluding entries for such individuals, or including them but making them invisible to all except the local Directory Service manager or other appropriate people.

Some information considered confidential will not be released, and additional attributes that may be considered for the next stage of the Service may be optional for the current Directory. In summary, each organization will have to strike a balance between standardization and flexibility in creating Directory entries.

The Right to Be in the Directory

The subjects of the data held in the Directory have well-defined legal rights, including the right of access to the information, the right to have inaccurate entries corrected, and the right to compensation for inaccuracy, loss, or unauthorized disclosure of information.

The terms and conditions that an employee accepts on taking up a post will (ideally, from the Directory viewpoint) include acceptance of a default directory entry established by management. A special clause will be included for the ex-directory case already discussed. On becoming an employee of any organization, an individual inevitably agrees to forego certain personal privacies and to accept restrictions. This is regarded as part of everyday life. The Directory is a larger step along the same road, enabled by increasingly sophisticated technologies. No doubt there will be others. But many subjects are likely to be quite happy to make their name and address available, and they should be given the opportunity to do so.

Concerns of the Individuals

The most obvious concern is one that could affect everyone. Unsolicited "junk mail" of all kinds pours through people's letterboxes at home and work daily. Some people like it, but many find it wasteful and intrusive. Many potential Directory subjects have expressed fears that they will be inundated with massive sales campaigns, requests for information, or abusive messages. Women could suffer particularly in this last respect, being more often the target for offensive messages.

A second concern for individuals is that of restricting access to the information in the Directory. Subjects will wish some information to be accessible to others in their own department, or organization, their own country perhaps, but not to anyone in the world.

Management Concerns

Probably the greatest concern expressed in an academic institution is management's reluctance to allow the development of comprehensive Directory data. Thus, some Directories will be incomplete and virtually useless. Management must be confident of two things: that there are adequate controls against serious invasions of the privacy of persons for whom they feel a responsibility; and that prevention of access by unauthorized persons to certain parts of the data is feasible.

Personnel managers are understandably worried about the possibility of sensitive information being widely available. This means that very secure access controls and authorization must be applied to any data they allow into the Directory. One solution, of course, is to simply not include the most sensitive data. Security of information can be treated as an access control issue and can be provided by the normal computer security features built into most systems. This is perhaps less of a problem than some concerns already discussed because a level of security, once defined, can be built into the service from the start and would not require frequent modification.

The decentralized nature of the Directory gives each organization complete control over its own information.

General Concerns

Higher management will be asked to release employee information to those who will be responsible for creating and maintaining a Directory Service. Employees may have conflicting feelings, divided between wanting as much useful information as possible included in the Directory but wishing to avoid the problems incurred by possible invasions of privacy. Both viewpoints are valid. Most staff would be justifiably concerned if they thought their administration departments were freely handing out sensitive information about them. There is also a more general concern relating to the "fair and lawful" collection of the information.

Although people who are subjects of Directory Service entries have a right to know what information is held about them, some administrators feel that no data should be included without first getting the specific permission, in writing, of the subjects. If such an agreement is reached, an enormous administrative effort will be required to issue and collect request forms. Some European countries have laws that insist this be done. The requirement is stringent, but proper administrative procedures will result in a slower rate of increase of information in these parts of the Directory. No judgment is being made here as to which is the "better" approach, as compliance is required with whatever law pertains. If each person who is to be the subject of an entry in the Directory is required to give active approval before the entry can be made, the natural inertia of most people will result in a much lower rate of participation than desired. This will seriously reduce the usefulness of the worldwide Directory.

The United Kingdom's law, as presently stated, allows the "data user" to use the information for a defined purpose, even if the subject complains. Data users are not legally obliged to tell the subjects before they include them, but in practice, someone setting up a Directory on behalf of their employer would be foolish to try to ignore existing restrictions. It is hoped this problem can be overcome by persuading the management of the reliability of the controls available in the Directory.

Maintaining Control of the Information

The ideal status for information in the Directory from the user's point of view is that all information relating to *other people* be fully accessible. The ideal status from the *subjects'* point of view is that information about *themselves* be available only to persons of benign intention. Since there is no way of knowing intention until it is too late, each organization must agree on a compromise on availability.

The decentralized nature of the Directory gives each organization complete control over its own information. It can set access controls in whatever way it wishes. The X.500 standard allows each organization to specify its own security policy. An organization can allow access to a variety of subsets of the information in its local Directory, in addition to allowing worldwide access to selected information, by including access control lists in the

database. There are two separate components: authorization, which is a method to specify, enforce, and maintain access rights to the information under its control; and authentication, which is a method to verify the identity of the users and the database holding the information and to verify the origin of information received.

Authorization must be given to the local manager of the information to modify all entries; the subjects may be allowed to modify all or part of their own entry, but not the entries of others. However, the organization may adopt the policy of holding the name and job title of every member of staff by default. Individuals could then be allowed to modify other parts of their entry, such as professional qualifications, but not the name and job title.

Some sections of the information may fall naturally into the category of "internal" information, access to which may be restricted to members of a particular department or subgroup within a department. For example, this restriction could be applied in a university to any information about undergraduates, where the class tutor may have access only in order to make class lists, while the matriculation office may be able to modify the entries as required. Outsiders would be granted no rights of access to this information.

Simple authentication is provided by the use of a password as proof of identify of the user. This is required before a subject is allowed access to his or her own information for modification. A higher level of authentication, involving cryptographic techniques, is required for information that needs a greater degree of security. Organizations will decide for themselves what level of security they wish to maintain.

Management and Maintenance

Once the initial Service is running, the importance of the update procedures cannot be overstated. The management of an organization may delegate the maintenance of the service to its computing division, but other organizational units must be prepared to supply regular updates of information, for example, details about new employees could be provided by the personnel department. It is essential that the maintenance be integrated into the everyday work of the responsible department if the Directory Service is to become maximally useful to the organization. In the early implementation period, when the variety of data is limited, procedures should be developed for the Directory to expand as more types of data are included in it.

Most organizations will have an administrator or manager for the Directory Service who is responsible for the accuracy of the information. Maintenance of the local Directory will be an integral part of normal administrative procedures within the organization, but an overall manager will still be required. A clear decision must be made as to who in the organization will provide update information for the Directory. If this is not done from the start, the reporting/update function may fall between two areas and not be done at all. This will put the organization in breach of any data protection laws, and in addition may render the service useless by its unreliability.

Accuracy of Information

Not only is it important that the information held be correct for the benefit of those trying to make use of it, but also there is likely to be a legal obligation that it be accurate. The UK Data Protection Officer is very concerned that data be accurate and not excessive for the purpose for which it is held. He or she is quite prepared to take enforcement action against those who exceed their registered requirements.

Conclusions

Establishing a Directory Service within an organization will involve a great deal of effort. It is essential that the confidence and cooperation of management is sought from the start of the project.

One of the first actions should be to ensure that Directory Service uses and advantages are made clear to the senior management. Management should be asked to agree on the inclusion of a minimum subset of information about each person which will make all entries useful to users. The reaction of management to these proposals varies greatly but is often on the cautious side. If pressed to reduce the initial amount of information entered, the fallback position could be to make optional fields available, which the subjects could choose to have entered. Adherence to a code of practice for data users will be helpful in building confidence by both management and staff in supporting a Directory Service.

Update procedures are of prime importance in making a Directory Service useful. To be effective, these procedures must be integrated into everyday

activities. For people to make full use of the Directory Service, they must feel confident that the data it contains is reliable and accurate.

The guidelines offered in this article are intended to assist organizations in planning and implementing a Directory Service. A successful X.500 Directory Service will be one of the most important tools for communication in the network environment and will fully repay the effort required to set it up.

Acknowledgments

I am happy to acknowledge the work of the UK Academic Pilot group chaired by Jim Craigie (JNT), the valuable advice given by the UK Senior Assistant Data protection Registrar, and the encouragement of my late Director, Allen McTernan.

Note

1. Eight recommendation areas make up the CCITT X.500/ISO 9594 Directory standard, Information processing systems—Open systems interconnection—The Directory—Parts 1-8. (1988):

- CCITT X.500/ISO 9594-1: *Information processing systems—Open systems interconnection—The Directory—Part 1: Overview of Concepts, Models and Services.*
- CCITT X.500/ISO 9594-2: *Information processing systems—Open systems interconnection—The Directory—Part 2: Models.*
- CCITT X.500/ISO 9594-3: *Information processing systems—Open systems interconnection—The Directory—Part 3: Abstract Service Definition.*
- CCITT X.500/ISO 9594-4: *Information processing systems—Open systems interconnection—The Directory—Part 4: Procedures for Distributed Operations.*
- CCITT X.500/ISO 9594-5: *Information processing systems—Open systems interconnection—The Directory—Part 5: Protocol Specifications.*
- CCITT X.500/ISO 9594-6: *Information processing systems—Open systems interconnection—The Directory—Part 6: Selected Attribute Types.*
- CCITT X.500/ISO 9594-7: *Information processing systems—Open systems interconnection—The Directory—Part 7: Selected Object Classes.*
- CCITT X.500/ISO 9594-8: *Information processing systems—Open systems interconnection—The Directory—Part 8: Authentication Framework.*

These standards can be obtained from the American National Standards Institute, Inc., 1430 Broadway, New York, NY 10018.

References

Cope, Patricia. (1991, September 30). Big users start moving to X.500 but face problems. *Network World* 8(30), 1 + 31-34, 45–46.

Dalton, Marion L. (1991). Does anybody have a map? Accessing Information in the Internet's virtual library. *Electronic Networking* 1(1), 31–39.

Dempsey, Lorcan. (1991). *Libraries, networks and OSI.* Westport, CT: Meckler (in association with the UK Office for Library Networking).

Planka, Daniela. (1990). Network directory services. *Library Hi Tech,* 8(4), 93–103.

Prospero: A Tool for Organizing Internet Resources

B. Clifford Neuman

Recent growth of the Internet has greatly increased the amount of information that is accessible and the number of resources that are available to users. To exploit this growth, it must be possible for users to find the information and resources they need. Existing techniques for organizing systems have evolved from those used on centralized systems, but these techniques are inadequate for organizing information on a global scale.

This article describes Prospero, a distributed file system based on the Virtual System Model. Prospero provides tools to help users organize Internet resources. These tools allow users to construct customized views of available resources, while taking advantage of the structure imposed by others. Prospero provides a framework that can tie together various indexing services producing the fabric on which resource discovery techniques can be applied.

The Internet contains a massive amount of information, but it is hard to use that information. There are several barriers to usability: it is difficult to identify the information of interest; it is difficult to keep track of this information once found; it is difficult to share information about what is available, or to collaboratively maintain such meta-information; and the information is often scattered across multiple file systems of different types, meaning that different mechanisms are needed to access it. Existing methods for organizing information have evolved from techniques used on centralized systems and are inadequate for organizing information on a global scale.

Users look for information in many ways. They consult libraries, journals, professional society publications, mailing lists, indexing services, and other users. While these sources of meta-information are useful, it is still necessary for users to identify the source that can answer their query. Prospero provides a framework within which such meta-information (which I will refer to as directories) can be made available to users, and it provides the tools to allow directories from multiple sources to be combined in useful ways.

Prospero lets users create customized views of a global file system. This customization plays an important role in organizing information since there are many communities of users, and they do not share the same interests. By supporting multiple views of the available information, one can improve the ease with which one finds information that is likely to be of interest, while keeping less useful information out of the way where the user is less likely to trip over it.

A prototype of Prospero is available and has been used to organize information on Internet sites world-wide. Prospero-based applications are used on more than 7,500 systems in 29 countries on six continents.

Organizing, Not Just Searching

There are four areas where work is needed to help users obtain the information they need: retrieval, indexing, search, and organization. A number of recent

***B. Clifford Neuman** is a computer scientist at the Information Sciences Institute of the University of Southern California. The work described in this article was begun while completing his Doctorate at the University of Washington. Neuman may be reached at USC/ISI, 4676 Admiralty Way, Marina del Rey, CA 90292-6695, USA. Telephone +1 (310) 822-1511, email bcn@isi.edu.*

This research was supported in part by the National Science Foundation (Grant No. CCR-8619663), the Washington Technology Center, Digital Equipment Corporation, and the Defense Advance Research Projects Agency under NASA Cooperative Agreement NCC-2-539.

The views and conclusions contained in this article are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any of the funding agencies.

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 30–37.

systems have addressed the first three areas, yet the fourth has been greatly ignored. Users require all four functions if they are to obtain the information they need. Without work on organization, the other functions become less useful as a system grows.

Some recently distributed file systems support a global name space. The Andrew File System (Howard et al., 1988) is an example. Such file systems provide for the retrieval of files worldwide, yet they do little to help the user find files of interest. Such file systems have directories near the root named after organizations, with the next level usually naming individual users. Files on particular topics are scattered across the leaves of the tree, where they are difficult to find.

Indexing can help users find information that is scattered across a distributed system. In attribute-based naming (Peterson, 1988), a name is resolved by querying a database of the attributes associated with local resources. Similarly, the Wide Area Information Server (WAIS) maintains a full-text index of a collection of documents, allowing users to search for documents by specifying words that appear in the full text (Kahle & Medlar, 1991). The Semantic File System (Gifford et al., 1991) provides another example of indexing by maintaining an index for all files on a collection of file servers. Distributed indexing (Danzig et al., 1991) provides an alternative approach to indexing widely distributed information. Indices are maintained by topic, and a topical index can request that future updates to other indexes be propagated if they match certain criteria. The indices in the systems described so far cover only a subset of the files that are available globally. It is still necessary for the user to find the correct server to query (selecting the index to be used).

Although it is possible to construct indices that cover large collections of files, it is necessary to trade detail and completeness for manageable size. For example, the Archie database (Emtage & Deutsch, 1992) indexes files from certain directories on major Internet FTP sites. The index, however, is based only on file names, not the file contents or other attributes. Completeness is also limited since only files available by anonymous FTP on selected sites are included. Another problem is that many queries return much more information than most users are prepared to deal with. In many cases, the large number of items found obscures the few that are really of interest.

When resources of interest to a user are distributed across multiple systems, and when the directory information needed to discover such resources is scattered across multiple indices, resource discovery techniques are needed to search for the desired information. Simplistic search strategies such as global broadcast or exhaustive depth-first search (as used by the Unix find command) are not suitable for large systems. Instead, search techniques should be based on browsing: looking at the information presently available and expanding the search in directions most likely to yield the desired results. Such browsing might include an interactive dialogue with the user (as is the case for directory browsers), it might be highly automated while accepting input from the user to narrow the search (as is done in Schwartz and Tsirigotis' (1991) resource discovery work), or once initiated it might run independently, returning the results to the user [knowbots (Kahn & Cerf, 1988) fall into this class].

Such search strategies are useful primarily when information is organized in such a way that programs and users can determine the appropriate direction in which to expand a search. One way to do this is to build a hierarchical directory service that can be used to find indexing services with information on various topics. Dalton (1991) discusses the possibility of using X.500 for this purpose. Such an approach works best when organizing a limited number of objects or when a single administrator can decide what is to appear in the upper levels of the name space.

The X.500 approach breaks down administratively, however, if used to organize fine-grained objects on a global scale. It is very difficult to gain agreement on what topics should appear near the top of the tree, and once topics are agreed on, there is disagreement on which resources should be included under each topic. This problem is apparent on Usenet, a worldwide distributed message service for disseminating messages on many topics. A significant share of the messages sent on Usenet discuss what messages are appropriate for particular newsgroups, whether new newsgroups should be created, and what they should be called. This clearly demonstrates the problem of reaching consensus on globally shared names.

Instead of supporting a single hierarchy for organizing information, it is possible to allow each user to organize information on his or her own. This customization is important for a number of reasons: it reduces the clutter that would otherwise be caused by resources in which the user has little interest; it allows users to define shorter names for frequently referenced resources; and it allows users to replace entire portions of the naming hierarchy with alternative views more appropriate for their particular

needs. User-centered naming also eliminates the need for consensus when deciding what should appear in the upper levels of the naming hierarchy. Each user can make that decision based on his or her own opinions.

Organizational mechanisms must make available directory information from many sources, including existing indexing schemes[1] and directory information specified by users. It should be possible for directory information from different sources to be combined in useful ways.

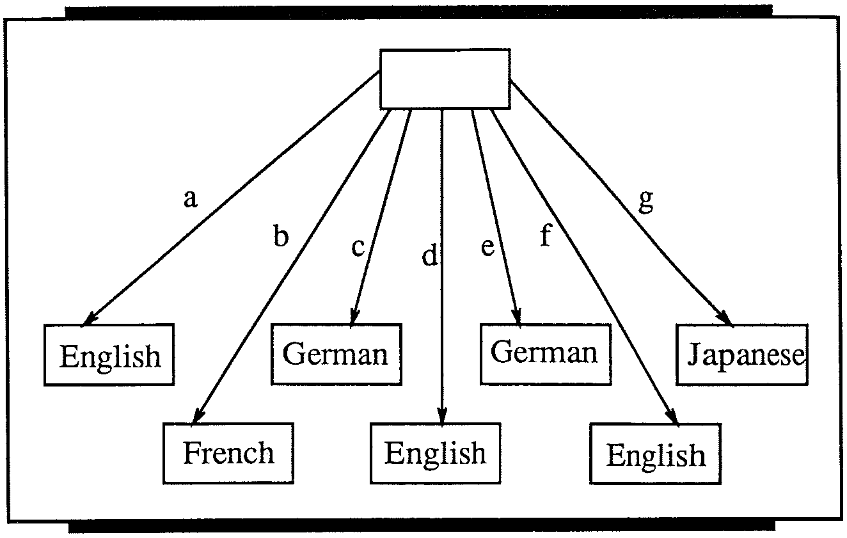


Figure 1. Directory before application of a filter

The Virtual System Model

The Virtual System Model (Neuman, 1992) provides a framework for organizing large systems within which users construct their own "virtual" systems by selecting objects and services that are available over the network; users then treat the selected resources as a single system, ignoring those resources that were not selected. The Prospero file system is a file system based on the Virtual System Model. By supporting a customized view of the system, information of interest to a user is prominently located near the center of the user's name space, while information that is not of interest is kept out of the way.

As users organize virtual systems for their own use, the structure imposed on the information can often be used by others. The Prospero naming network forms a generalized directed graph. A user's name space appears hierarchical and corresponds to the names seen by the user starting from a particular node in the graph, the root of the name space. If a user finds an object or a directory of interest, the user can add a link that will make the object more prominent. When a user creates a directory with links to objects on particular topics, others can (if authorized) view that directory and include it in their own virtual systems, thus benefiting from the organization imposed by the first user.

Indexing services are made available through Prospero by treating the results of a query as a virtual directory. Users can add links to the directories corresponding to particular indices, and even to directories that correspond to queries executed upon those indices.

Two features of Prospero allow new views of information to be derived from meta-information that already exists. If a union link is included in a directory, the contents of the directory that is the target of the link appear to be included in the directory containing the link. This allows a directory to incorporate directory information from other sources. When the original source changes, the changes will also be reflected in the directory incorporating that information.

When constructing views, users can also associate functions (filters) with links that allow the creation of derived views from views that already exist. For example, in Figure 1 files are named with the labels *a* though *g*. Associated with each file is an attribute list, one attribute of which is the language in which the text was written. The value of the language attribute is shown in the box representing the file. By attaching the **distribute()** filter to the directory link, a derived view is created within which the files appear to be distributed across subdirectories according to the value of the language attribute. The derived view is shown in Figure 2.

A filter can be an arbitrary program that takes a representation of a directory as an argument and returns the same. It can add links to a directory, delete links, change the names of links, and even define new filters that are to be applied when traversing links deeper in the hierarchy. As arbitrary programs, filters can access any information needed to perform their function. Typically, this information includes attributes of files and the contents of other

directories, but it might involve reading files or performing database queries. Although users can write their own filters, it is expected that most will use the set already defined for them.

Organizing Information with Prospero

The Virtual System Model allows information to be organized in many ways, and many parties will play a role in doing so. Among the entities that will organize information will be individuals, professional societies, libraries, governments, commercial indexing services, or any collection of individuals sharing a common interest. An important feature of the model is that the same information can be organized in multiple ways.

The individual in the best position to organize the papers written by a particular author is that author. With Prospero, an author can maintain a directory referencing his or her own work, or at least that work which others should find. The incentive for doing so is visibility. The ease with which others can find one's writings affects the likelihood that those writings will be used. By maintaining one's own index of papers, one can also add cross-references to more recent work as it is completed.

The usefulness of such a directory is greatly enhanced when it is itself referenced from a higher level directory of authors. Such directories are maintained today in library card catalogs and in reader's guides to the literature, but the job of maintaining such directories is greatly simplified when implemented using Prospero; the maintainer of the higher level index would only have to update the directory when new authors are added. Once added, it is up to the authors themselves, or to individuals maintaining directories on behalf of the authors, to keep the list of the author's publications current.

Organizations like the ACM and the IEEE might each maintain a directory of topics in computer science and designate experts in each area to maintain the directory on that topic. Organizations in other fields, for example, the American Medical Association, might do the same. The custodians of particular topics could add references to worthwhile items as they are discovered. In cases where certain well-crafted queries on automatically maintained databases yield useful results, those queries can be encoded in filters, and the result added to the collection of topics as a virtual directory. Libraries could then maintain directories of general fields such as computer science, chemistry, and literature with links to the directories maintained by various organizations.

Users will build their own hierarchies of files by creating directories, subdirectories, and files of their own and by adding links to files, directories, and subdirectories created by others. Files that are frequently accessed by a user will probably have short names while names will be longer for objects of less interest. Because directories of other users will be accessible from the user's virtual system, the virtual system will probably contain files that a user has never accessed and might not even know about. These files, however, will be deep in the user's hierarchy.

If individuals do not like the way information is organized, they can organize it themselves, or they can find different experts whose views more closely match their own. They can completely customize their own name space so that their alternative view is used instead of the more accepted view. In fact, which view is the accepted view becomes more a matter of whose views more people adopt, rather than whose view is officially sanctioned.

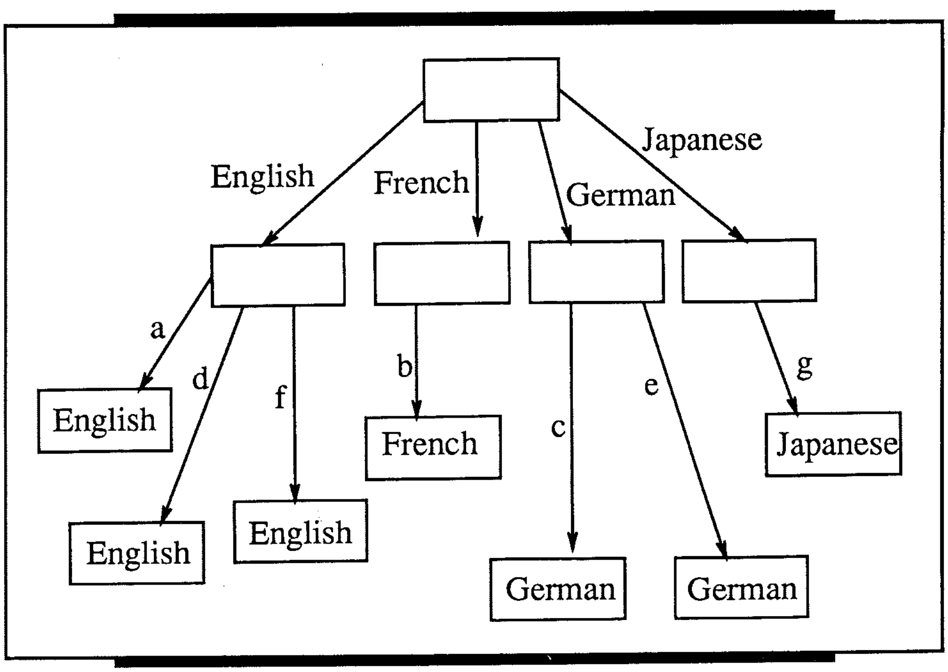


Figure 2. Directory with distribute() applied

Over time, multiple communities of users will evolve. It is expected that the members of each community will have similarly structured name spaces, but name spaces may vary widely across different communities of users. For example, members of the computer science community might organize virtual systems in one way, while members of the medical community might think of the world in a completely different manner.

Searching for Information

Once information has been organized, users can look for it in many ways. A user looking for a paper on heterogeneous computer systems by a particular author might find the paper in a directory maintained by that author. A user who did not know any of the authors might find the same paper in a directory of papers on distributed computing. Of course, just knowing that the information of interest exists in a published paper can be a big help; many times a user will not even know that.

Today, if something is available that is of interest, it is often found through directories such as the phone book or yellow pages, through reading newspapers and other periodicals, or by word of mouth. In the research community, these sources of information are supplemented by technical papers, electronic mail, and mailing lists. It is likely that these methods will continue to find significant use even once other mechanisms are in place. The Virtual System Model allows much of the information that is useful for finding objects, but which to date could only be obtained by external means (such as asking the author of a paper), to be included as part of the file system. The Prospero file system can then be used as the matrix through which users can navigate to find the desired information.

One way that information can be found using the Prospero directory service is through browsing. An individual interested in a particular topic can connect to the virtual system of someone else who is known to be interested in that topic.[2] The user could then look through those virtual systems for documents or files of interest. Of course, users would only see those files that the owner of the virtual system has authorized them to see.

Browsing is considerably more likely to be effective using Prospero than with traditional file systems. Prospero encourages users to make their own links to the files in which they have an interest. As such, interesting files are likely to appear in the hierarchies of many people, thus increasing the likelihood that the files will be found by browsing.

The Prospero directory service also provides the fabric on which resource discovery methods might operate. The Prospero server makes directory structures from existing systems part of that fabric, yet users can add their own links to augment the existing structure. Knowbots could navigate through the fabric and might themselves augment the existing structure by adding links to objects that they find. This augmentation of the naming network might provide both a method for a Knowbot to communicate its results back to its initiator, as well as a method through which knowbots can interact with each other.

Experience

A prototype of Prospero has been available since December 1990.[3] The prototype allows users to construct virtual systems and to navigate through them. In addition to the basic release, there are several standalone applications that rely on Prospero to retrieve directory information from indexing services.

Programs linked with the Prospero compatibility library are able to specify file names relative to the active virtual system when opening files. Prospero is a heterogeneous file system; instead of providing its own methods for accessing files, it relies on multiple underlying methods. The prototype presently supports Sun's Network File System, the Andrew File System, and the File Transfer Protocol (FTP). For FTP, the file is automatically retrieved, and the locally cached copy is then opened.

As distributed, a user's virtual system starts out with links to directories organizing information of various kinds in several ways. Figure 3 shows a sample session with Prospero. Users find information by moving from directory to directory in much the same manner as they would in a traditional file system. Users do not need to know where the information is physically stored. In fact, the files and directories shown in the example are scattered across the Internet. At any point, a user can access files in a virtual system as if they were stored on his or her local system.

In the example, the user connects to the root directory and lists it using the `ls` command. The result shows the categories of information included in the virtual system. The information includes online copies of papers (in the papers directory), archives of Internet and Usenet mailing lists (in the mailing-list and newsgroups directories), releases of software

```
Script started on Wed Jan 29 21:02:50 1992
% cd /
% ls
afs                info               papers
databases          lib                projects
documents          mailing-lists      releases
guest              newsgroups         sites
% cd papers
% ls
authors             conferences        subjects
bibliographies      journals           technical-reports
% cd technical-reports
% ls
Berkeley     IAState        OregonSt      UCalgary    UWashington
BostonU      MIT            Purdue        UColorado   Virginia
Chorus       NYU            Rochester     UFlorida    WashingtonU
Columbia     NatInstHealth  Toronto       UKentucky
Digital      OregonGrad     UCSantaCruz   UMichigan
% ls UCSantaCruz
crl
% ls UCSantaCruz/crl
ABSTRACTS.1988-89             ucsc-crl-91-01.ps.Z
ABSTRACTS.1990                ucsc-crl-91-02.part1.ps.Z
ABSTRACTS.1991                ucsc-crl-91-02.part2.ps.Z
ABSTRACTS.1992                ucsc-crl-91-02.ps.Z
INDEX                         ucsc-crl-91-03.ps.Z
ucsc-crl-88-28.ps.Z           ucsc-crl-91-06.ps.Z
...
% ls UWashington
cs   cse
%
% ls UWashington/cs
1991           INDEX           PRE-1991
1992           OVERALL-INDEX   README
% cd /papers
% ls
authors             conferences        subjects
bibliographies      journals           technical-reports
% ls journals
acm-sigcomm-ccr  ieee-tcos-nl
% ls journals/ieee-tcos-nl
app-form.ps.Z  v5n1           v5n3
cfp            v5n2           v5n4
% ls journals/acm-sigcomm-ccr
application.ps  jan89           jul90           sigcomm90-reg.ps
apr89           jan90           oct88
apr90           jan91           oct89
apr91           jul89           sigcomm90-prog.ps
% vls journals
   acm-sigcomm-ccr       NNSC.NSF.NET      /usr/ftp/CCR
   ieee-tcos-nl          FTP.CSE.UCSC.EDU /home/ftp/pub/tcos
%
script done on Wed Jan 29 21:06:53 1992
```

Figure 3. Sample session

packages (in the releases directory), and the contents of prominent Internet archive sites (in the sites directory). Files of interest can appear under more than one directory. For example, a paper that is available from a prominent archive site might also be listed under the papers directory.

Next, the user connects to the papers directory, lists it, and finds the available papers further categorized as conference papers, journal papers, or technical reports. The technical report directory is broken down by organization and by department within the organization. The journals directory is organized by the journal in which a paper appears, and the two journals that are shown are futher organized by issue. Use of the `vls` command shows where a file or directory is physically stored, demonstrating the fact that the files are scattered across the Internet (IEEE TC/OS Newsletter on FTP.CSE.UCSC.EDU and Computer Communications Review on NNSC.NSF.NET.) Though not shown in the example, papers are also organized by author and subject in other directories from the same virtual system.

It is important to note that the example shows only part of the information available through Prospero, and that it shows a typical way that the information is organized. Individuals can organize their own virtual systems differently.

One of the most frequently used directories in Prospero is that representing the Archie database, developed at McGill University (Emtage & Deutsch, 1992). That directory includes subdirectories organizing files according to the last components of their file names. For example, the subdirectory prosp contains references to the files available by Anonymous FTP whose names include the string *prosp.* Among the matches would be files related to Prospero. The contents of each subdirectory are equivalent to what would result from running the Unix `find` command with appropriate arguments over all the major archive sites on the Internet (if it were even possible to do so). The subdirectories do not exist individually but are instead created when referenced by querying the Archie database. The use of Archie through Prospero has been so successful that the Archie group has adopted Prospero as the preferred method for remote access to the Archie database.

To provide the benefits of Prospero to users who have not installed it on their systems, Steve Cliffe of the Australian Academic and Research Network (AARNet) Archive Working Group has added Prospero support to one of their FTP servers. As well as making files available from the physical file system, the modified FTP server makes files available from a virtual file system. When a retrieval request is received, the FTP server locates the file using Prospero and checks to see if a copy of the file is available locally. Using Prospero to check the last modified time of the authoritative copy, the FTP server checks that the local copy is current. If a current copy does not exist locally, the server retrieves and caches a copy of the file. The local copy is then returned to the client.

Future Plans

Prospero is an evolving system. We are continuing to work closely with the Archie group to make additional databases available. Immediate plans for the future also involve integrating Prospero with additional indexing services including WAIS (Kahle & Medlar, 1991), and once they are deployed, semantic file systems (Gifford et al., 1991) and distributed indices (Danzig et al., 1991). This will be accomplished by allowing a Prospero server to make meta-information from these databases available using the Prospero protocol.

In many respects, the goals of Prospero are similar to those of Hypertext systems such as World Wide Web (Berners-Lee et al., 1992). We hope to make information from that system available through Prospero.

We will be adding additional methods for retrieval of data. This will be of use when integrating WAIS and Prospero since much of the data indexed by WAIS is retrievable only with Z39.50. In addition to adding real-time access methods, we will be adding several off-line methods. For files that are accessible only by electronic mail, an e-mail method will be added that will automatically request the file on the user's behalf, allowing references to such files to be organized together with other files.

We also plan to add support for publications that are available only on paper. Indices for such information can be made available by running a Prospero server over a bibliographic database. The references would indicate the information needed to obtain a copy of the document, either an ISBN number or perhaps the shelf location in the local library.

Concluding Remarks

The Virtual System Model provides a powerful framework within which information can be organized. Prospero makes that framework available for

organizing information on the Internet. By themselves, neither the model nor the prototype helps users find information of interest. Their contributions are in encouraging and enabling users to organize information in ways that make it easier to find things.

Professional societies, libraries, governments, commercial indexing services, and others will play important roles in organizing the information available from future systems. The Virtual System Model allows such service providers to build on each other's work, eliminating duplicated effort, and it allows users to construct views of the information provided by these services which better meet their own requirements. The real contribution of this work will depend on the extent to which the model is adopted by these service providers and how it is used in future systems.

Acknowledgments

Ed Lazowska provided valuable guidance throughout this work. Discussions with John Zahorjan, Hank Levy, and Alfred Spector helped me refine the ideas that ultimately led to the development of Prospero. Celeste Anderson, Ben Britt, Steve Cliffe, Peter Danzig, Peter Deutsch, Alan Emtage, Deborah Estrin, Dennis Hollingworth, and Charles McClure provided comments on earlier drafts of this paper.

Notes

1. In fact, indexing is itself a method for organizing information, although it is typically applied to only a subset of the information available.

2. The directories and files that a user maintains will be owned by that user. Parts of a user's hierarchy, however, may be owned by other users. Access control information is maintained along with each file or directory, and with each directory link. This information determines who is allowed to read the file or search the directory. It is expected that users will make parts of their hierarchies accessible to others, but how much is to be made available will be decided by the individual.

3. For information on obtaining the release please send a message to info-prospero@isi.edu.

References

Berners-Lee, Tim, Cailliau, Robert, Groff, Jean Francois, & Pollermann, Bernd. (1992). World-wide web: The information universe. *Electronic Networking: Research, Application, and Policy*, 2(1), 53–59.

Dalton, Marian L. (1991). Does anybody have a map? Accessing information in the Internet's virtual library. *Electronic Networking: Research, Application, and Policy*, 1(1), 31–39.

Danzig, Peter B., Ahn, Jongsuk, Noll, John, & Obraczka, Katia. (1991). Distributed Indexing: A scalable mechanism for distributed information retrieval. *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 220–229). New York: ACM Press.

Emtage, Alan, & Deutsch, Peter. (1992). Archie: An electronic directory service for the Internet. *In Proceedings of the Winter 1992 USENIX Conference* (pp. 93–110). Berkeley, CA: USENIX Association [2560 Nonth St., Suite 215, Berkeley, CA 94710].

Gifford, David K., Jouvelot, Pierre, Sheldon, Mark A., & O'Toole, James W. Jr. (1991). Semantic File Systems. *Proceedings of the 13th ACM Symposium on Operating Systems Principles* (pp. 16–25). New York: ACM Press.

Howard, John H., Kazar, Michael L., Menees, Sherri G., Nichols, David A., Satyanarayanan, M., Sidebotham, Robert N., & West, Michael J. (1988). Scale and performance in a distributed file system. *ACM Transactions on Computer Systems*, 6(1), 51–81.

Kahle, Brewster, & Medlar, Art. (1991). *An information system for corporate users: Wide area information systems.* Technical Report TMC-199. Menlo Park, CA: Thinking Machines Corporation.

Kahn, Robert E., & Cerf, Vinton G. (1988). *The Digital Library Project; Volume 1: The world of knowbots* (Draft). Available from Corporation for National Research Initiatives, 1895 Preston White Dr., Suite 100, Reston, VA 22091.

Neuman, B. Clifford. (1992). *The Virtual System Model: A scalable approach to organizing large systems.* Doctoral dissertation (in preparation). Seattle,WA: University of Washington: Department of Computer Science and Engineering.

Peterson, Larry L. (1988). The Profile naming service. *ACM Transactions on Computer Systems*, 6(4), 341–364.

Schwartz, Michael F., & Tsirigotis, P. G. (1991). Experience with a semantically cognizant Internet White Pages Directory Tool. *Journal of Internetworking: Research and Experience*, 2(1), 25–50.

HYTELNET as Software for Accessing the Internet: A Personal Perspective on the Development of HYTELNET

Peter Scott

The "community" of computers commonly referred to as the Internet contains vast amounts of information useful to librarians, scholars, networkers, businesspeople, professionals, and the general public. This information comprises online public-access catalogs, full-text databases, campuswide information systems, bulletin boards, and other types of knowledge bases. Until recently, discovering what is available has been a painful chore for the user. Paper directories exist, but they are out of date as soon as they are published, and they are cumbersome to update. The HYTELNET software, which gives a user the login addresses and passwords to every known remote site on the Internet, has made the process of finding sources easier. HYTELNET guides a user, with hypertext jumps, through the maze of information sources. This article explains how the program operates, what it comprises, and how it can be updated.

Internet Resources and HYTELNET

There is an enormous amount of useful and interesting information residing on that vast, and somewhat intimidating, resource known as the Internet. The Internet is not a network in its own right. Rather, it is the name given to the 5,000 or so networks, situated in about forty countries, which comprise a "community" of roughly half a million computers serving millions of users. Given its size, geography, and lack of any truly formal organizational standards, it is little wonder that a new user will feel frustration, confusion, and despair when attempting to locate its information resources. Paper directories of resources are useful, but they lose their currency very quickly if not updated regularly. They also may not be all-inclusive. Certain kinds of resources may be omitted owing to the compilers' lack of interest, time, or knowledge.

In order to alleviate the pain of determining the availability and location of the Internet's electronic information, HYTELNET was developed. It was designed with one goal in mind: to make access to Internet resources as easy as possible for both new and experienced networkers, ensuring that the information being presented was timely, accurate, and understandable. It was the first software package for personal computer users that attempted to bring some order to the chaos of remote Internet login.

HYTELNET is a utility, developed in late 1990, that allows an IBM-PC user to gain almost instant access to all known sites on the Internet. It is an acronym for *HY*pertext browser for *TELNET*- accessible sites. Sites accessible with Telnet include hundreds of online public-access library catalogs, library bulletin boards, campuswide information systems, Free-Nets, full-text databases, "electronic" books, network information centers, and many other useful services scattered around the globe. HYTELNET also includes a glossary of Internet terms and a file containing instructions on the use of the telnet program itself. Instructions for retrieving HYTELNET are provided in the Appendix.

__Peter Scott__ <scott@sklib.usask.ca> is the Small-Systems Manager, University of Saskatchewan Libraries, Canada. He received his Bachelor of Arts degree from the Middlesex Polytechnic, England, in 1973. He has held many positions in the British book trade and various libraries. His major professional pursuits include teaching users how to access Internet resources and designing hypertext utilities. Scott may be reached at the University of Saskatchewan Libraries, College Drive, Saskatoon, Saskatchewan, Canada, 27N 0W0; (306) 966-5920.

The Pre-HYTELNET Situation: Paper Directories

In October 1990 our library VAX computer had the Telnet and FTP programs added. These programs allowed us to make connections to remote sites, so that we were able to login to other computers around the world in order to perform various types of on-line searching and file transfers. We already had the MAIL program, and I had recently taken out

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 38–44. © 1992 Meckler

subscriptions to a few electronic journals and conferences, including PACS-L, LIBREF-L, and CWIS-L.

One of the more interesting debates on PACS-L centered around two paper lists of Internet-accessible sites, commonly referred to as the Art St. George (1990) and Billy Barron (1991) lists. Both lists contained descriptions of and login procedures for, in most cases, library on-line catalogs. The debate centered around the notion that one of them was the most authoritative and complete, and that the other was a mere copy, simply reformatted. There was no mention of an electronic directory in the debate. That struck me as curious, since both lists were carrying information concerning an electronic procedure.

The paper lists were dutifully downloaded and used as guides to help us connect, using the Telnet program on our VAX to access remote sites. Connections were made to the Princeton and Harvard library catalogs, the Colorado Alliance of Research Libraries (CARL), and libraries in Germany, Mexico, and Australia.

A number of questions arose regarding the lists. Being paper lists, they would probably be out of date as soon as they were announced. So, how would they be updated to remain useful? It seemed likely that more sites would be added on an almost daily basis. Would a user want to keep downloading new paper copies and discarding the old ones, just because some new sites had been added? A user might want to connect to the remote sites from any number of computers. Did that mean that he would need to have the paper lists on his person at any given time?

I had previously compiled a hypertext utility for helping people understand all the intricacies of VAX MAIL. It is called HYPERVAX, (Scott, 1990) and the HYPERREZ (Larson, 1989) software from Maxthink was used to compile it. I decided to use the same software to compile a utility that would take the place of the paper directories and that would be available to a personal computer user with the touch of two keys.

The list written by Billy Barron of the University of North Texas looked like the better of the two lists for producing a hypertext equivalent. Entitled "UNT's Accessing On-Line Bibliographic Databases," the list is arranged alphabetically by site. Each entry is formatted in a standard fashion, containing full login, password, and logout procedures. Also listed, when available, is the software a particular site uses for cataloging its collection, for example, Geac, NOTIS, or BUCAT.

Developing the Software

The following discussion describes, in some detail, how Barron's paper list was manipulated to create the first version of HYTELNET. First, the file was loaded into the text-editor, QEdit, ("QEdit Advanced v2.15," 1991) a very fast, powerful, and easy to use program, ideal for creating hypertext files.

The next step was to determine how the information was to be arranged so that files with sensible hypertext links could be created. The beauty of HYPERREZ, the driver, is that it allows the linking of pure text files with hypertext jumps. A jump is a word surrounded by pointed brackets, which, when pushed, links to the file with that word as its filename.

After loading the program into the computer's memory, the driver calls up the file called START.TXT (see Figure 1). This, again, is a pure text file and can contain any ASCII characters. START.TXT, as its name implies, is the starting point for all subsequent jumps.

The START.TXT file for HYTELNET contains a link to a file named WHATIS (see Figure 2) which briefly describes the purpose of the program. It is the first highlighted link. Pushing the right arrow key makes the driver jump to it. Hitting the left arrow key returns to START.TXT.

```
                 Welcome to HYTELNET  version 5.0

                  What is HYTELNET?   <WHATIS>
Telnet-accessible library catalogs   <SITES1>
     Other telnet-accessible sites   <SITES2>
                 Internet Glossary   <GLOSSARY>
                Cataloging systems   <SYS000>
              Understanding Telnet   <TELNET>
               Key-stroke commands   <HELP.TXT>

           HYTELNET 5.0 was written by Peter Scott,
 U of Saskatchewan Libraries, Saskatoon, Sask, Canada.  1992
```

Figure 1. The START.TXT file

```
HYTELNET...... Program Description.....

♦ HYTELNET is designed to assist you in reaching all of the
           INTERNET-accessible libraries, Freenets, CWISs,
           Library BBSs, & other information sites by Telnet.
♦ HYTELNET is designed specifically for users who access
           Telnet via a modem or the ethernet from an
           IBM compatible personal computer.
♦ HYTELNET, when loaded, is memory-resident. Once loaded hit
           Control + Backspace to activate the program. To
           leave the program temporarily hit ESC. To remove
           from memory hit ALT-T while in the program.

♦ For information on customizing the program see <CUSTOM>
♦ For accessible Library on-line catalogs see <SITES1>
♦ For other information sites see <SITES2>

For extra information on loading the program and how to
      contact the author go to the <READ.ME> file
```

Figure 2. The WHATIS file

The original text file was broken into further files, each containing information on only one site. Each file was given a unique name starting with an abbreviation for a particular country or area, for example CN for Canada, AT for Australia, US for the United States. That being done, the next job was to create files that listed the names of the sites belonging to their particular country. These files, in turn, had to be linked to a file that contained the names of all the countries. This file is called SITES1 (see Figure 3). The SITES1 file is the second link on the file START.TXT.

Hypertext links can be placed in any file, as long as they lead somewhere. Links were also created for the various cataloging systems mentioned in the descriptions of the libraries (see Figure 4). These files are basically mini-help files for understanding how to search in a particular system, such as GEAC and DRA.

Creating hypertext utilities can be quite time consuming and frustrating, depending on the information being formatted. Fortunately, Billy Barron had created the perfect document for such purposes. Thus, the job at hand was to design a fairly complicated linking system that would not distract or sidetrack the user.

Other files that were felt to be necessary were created for HYTELNET. For instance, a file containing very brief instructions on how to use the arrow keys to make jumps from file to file; a READ.ME file explaining the purpose of the program, credits, and contact addresses; and a file called CUSTOM that explains how to edit a site file when necessary.

HYTELNET is designed to be a terminate-and-stay-resident program. In other words, it is loaded into memory and invoked when the user needs it. This makes it a perfect complement to a communications software package. For example, suppose the user were connected to a campus mainframe and decided that a visit to another site's catalog was required. The control and backspace keys are pressed, and HYTELNET pops up ready for browsing. The arrow keys allow for quick and efficient searching. The information is found, and the program is returned to memory while the user performs a Telnet search. Once connected the user can recall the program any time by reinvoking it.

Version 1.0 Released Early 1991

The first version of HYTELNET was released at the beginning of 1991 and was greeted with a positive response. The comments received indicated that many Internet users found it to be a useful and somewhat overdue utility. Most users were running the program from their personal computers, but it was becoming evident, based on the correspondence received, that computer wizards were beginning to adapt it for their own special needs. One such user was Richard Duggan, at the University of Delaware, who adapted the text files so that they would run on a Windows program he designed called CATALIST (Duggan, 1990).

As I pursued my interest in the Internet, I began to discover other types of sites to which a Telnet connection could be made. These included library

```
Library Catalogs arranged by country

        <AT000>    Australia
        <CN000>    Canada
        <FI000>    Finland
        <GE000>    Germany
        <IR000>    Ireland
        <IS000>    Israel
        <MX000>    Mexico
        <NE000>    Netherlands
        <NZ000>    New Zealand
        <ES000>    Spain
        <SW000>    Sweden
        <SZ000>    Switzerland
        <UK000>    United Kingdom
        <US000>    United States
```

Figure 3. The SITES1 file

bulletin boards, Campus-wide Information Systems, Free-Nets, and many different types of specialty services (see Figure 5). Information on these systems was briefly touched upon in the St. George list (1990), but it seemed that their full inclusion in HYTELNET was mandatory. So what began as an electronic guide to library on-line catalogs was rapidly becoming a complete directory of the Internet. In only a few weeks after the release of Version One, much new information had been discovered, including a file containing access instructions for all the United Kingdom libraries on the JANET network, as well as many more Canadian and Australian libraries. Version 2.0 was then compiled and distributed to the Internet community.

```
                          Using DRA Atlas

Author searches:          To search for a particular author, use the a=
                          search command followed by the author's name.
                          Example:   a=Haley Alex

Subject  searches:        To search for a particular subject, use the s=
                          search command followed by the subject.
                          Example:   s=Stars

Title searches:           To search for a particular title, use the t=
                          search command followed by the title.
                          Example:   t=Winds of War

Call Number search:       To search for a particular call number, use the
                          c= search command.
                          Example:   c=tr897.5

ISBN search:              To search for a ISBN, use the i= command.
                          Example:   i=1558511431

ISSN search:              To search for a ISSN, use the n= command.
                          Example:   n=0010-0285

LCCN search:              To search for a LCCN, use the l= command.
                          Example:   l=90012345

Music Publishers search:    To search for a Music Publishers #, use the
                        r= command.
                          Example:   r=CD 80096 telarc

Keyword search:           Type k.  Some DRA sites use the Z39.58 standard
                          for the keyword search.  See the section on
                          "Using Z39.58".         <Z39.58>

Help:                     Type ??.
```

Figure 4. The DRA help file

The Usenet newsgroups were also a mine of information, and, by reading the messages related to Telnet login, even more sites were found for inclusion in the utility. It was not just information that was discovered. Contact was made with many Internet users who were willing to share their own remote-access experiences. In order to take advantage of their knowledge and to offer information on the tricks of the Internet trade, a mailing list was created on our VAX, called LIB_HYTELNET (1990). Currently, about 300 members in nine countries who freely share knowledge and co-operate in finding information on new sites. As a result, the users themselves assist in creating new versions of the software and are free to adapt it in any way they choose.

```
          Other Telnet-accessible resources

<ARC000>   Archie: Archive Server Listing Service
<CWI000>   Campus-wide Information systems
<FREE000>  FREE-NET systems
<FUL000>   Full-Text Databases and Bibliographies
<LIBB000>  Library Bulletin Boards
<NAS000>   NASA databases
<NET000>   Network Information Services
<WAI000>   Wide Area Information Servers

<OTH000>   Miscellaneous resources
```

Figure 5. The SITES2 file

Capturing Information from Remote Sites and Creating Directories

Information found on the remote sites needs to be captured to a file if it is to be of any use. One of the best telecommunication software programs for remote capture is TELEMATE (Hu, 1992).

TELEMATE is a Canadian-designed communication program, similar to such programs as PRO-COMM and TELIX, but with one essential difference: it allows multitasking. Simply put, multitasking allows a user to perform various functions simultaneously. When live, TELEMATE records everything that appears on the terminal screen into a backscroll screen, thus allowing information to be reviewed at any time. It also has the ability to log sessions to a file. Its text editor can be invoked at any time, and the program allows for information to be copied from the backscroll screen into the editor and to be saved as a pure text file. When logging into a new Telnet-accessible site, terminal information can be captured, copied to the ed-

itor, and unwanted characters erased, saved, and finally mailed as a new file to all the members of LIB_HYTELNET. They in turn can add the new file to their own version of HYTELNET so that it is completely current.

If HYTELNET is not running, but a login to a remote site is needed, the particular file needed for login instructions can be loaded from a hard disk and placed in the View window in TELEMATE. All cut/copy/paste functions are available in the View window. This is particularly useful when a member of the messaging group reports some new information on a site. Only one file rather than the whole program needs to be updated.

When a new site is discovered, it is sometimes necessary to update more than one file. For example, suppose that the University of Alberta has made its catalog available on the network. Not only will its own file be created but also the file listing all the Canadian libraries will need updating, as well as the file listing the type of cataloging software being used. However, this takes very little time to do, given that all the files in HYTELNET are discrete, small, and easy to edit. To attempt that kind of updating with a paper file would be a nightmare!

HYPERREZ was chosen as the software to create the hypertext utilities because it is very easy to understand and use. The time taken by a novice user to learn how to navigate through the links is minimal, and the system is to some extent intuitive. It is ideal for creating hypertext directories. Not only have HYPERVAX and HYTELNET been created, but also Diane Kovacs' "Discussion List for Academics" was transformed into a utility called HYDIRECT, a new edition of which will be released later this year. For DOS users, HYDOS (Scott, 1991a) was compiled, giving instant access to all DOS commands. Also available is a utility that describes the commands for using Envoy100, with which most Canadian librarians are familiar. It is named, appropriately, HYENVOY (Scott, 1991b). Work continues on the compilation of utilities for TELEMATE, as well as on a browser for the new version of Kermit.

HYPERREZ is designed as a terminate-and-stay-resident program, which uses very little memory. HYTELNET, TELEMATE, QEdit, and a file manager can all run together in 640k of RAM. Apart from the driver itself, HYPERREZ allows pure ASCII files to be linked. Since there is nothing proprietary about ASCII, it is obvious that file editing becomes very easy. Paper directories present information in a linear fashion. Conversely, hypertext allows a reader to jump very quickly to a topic of interest and to make deeper jumps when required. The trick is to be able to return to a starting place, quickly and with little effort. HYPERREZ allows this.

When you obtain any of the hypertext utilities mentioned above, the driver, of course, comes with the package. All that is needed to use the program are a couple of essential files, which can easily be edited. A user of HYPERREZ can create many useful utilities, such as a library collections directory, a hypertext campuswide information system, a directory of library staff and hours, and an information package for new library users.

HYTELNET 5.0

HYTELNET is now in its fifth version and has been completely re-designed (Scott, 1992). Previously, all files were housed in one directory. This arrangement tended to slow down the hypertext jumps and to make the editing of the files very cumbersome. Now each distinct group of files resides in its own subdirectory, that is, all the Canadian libraries are filed under the CN0 subdirectory. When HYPERREZ searches for a file to display it looks for the first three characters in a subdirectory name, then at the appropriate file, and next it loads it. So, once again, a sensible arrangement of information files is essential for speed and efficiency.

Many Internet users have adapted the information contained in HYTELNET to allow it to run on platforms other than MS-DOS. Earl Fogel of the University of Saskatchewan Computer Services Department has created a UNIX version of HYTELNET (Fogel, 1992) that runs on the VAX mainframe. In order to gain access to Earl's version, a user issues the command: telnet ocdc.usask.ca, logging in with the username `hytelnet'. As with the PC version, hypertext jumps can be made very quickly with the arrow keys. This version is not memory-resident, however. In order to make a Telnet connection to a remote site, the user merely has to hit the ENTER key on a SITE screen when prompted. The source code for the UNIX version is available on the Internet.

Billy Barron has adapted the files in HYTELNET to be searched by the Wide Area Information Server (WAIS) at Thinking Machines Corporation in California (Barron, 1992). In order to search the database, a user will select HYTELNET from the directory of servers and choose a keyword or words as a query. The results will be retrieved and displayed almost immediately. This technology is in its infancy but can certainly be regarded as "the next big thing" in information retrieval. HYTELNET information is also

available on various "Internet Gophers" (1992). An "Internet Gopher" is an information distribution system that allows browsing of an information hierarchy.

As far as my own plans are concerned, I will continue to maintain the updating of the information in HYTELNET, keeping users abreast of any new and interesting sites that become available. A stand-alone version will be issued in the next couple of months. This version will allow a user to call up an editor and make any necessary changes to a file, print any file, search a file by keyword, and build a glossary of useful terms which can be accessed with one keystroke.

Full-time Internet Indexers Needed

Currently, Internet information directories, whether produced on paper or electronically, are compiled and maintained by interested amateurs. The body of work that has thus far been created is impressive and useful, but it could be argued that the time has come to make a permanent and full-time group of individuals responsible for developing such directories. The Internet grows daily. New resources are continually being made available to the community of users. HYTELNET, Wide-Area Information Servers, and the Gophers are merely the first steps toward creating a truly global information source. The issue of user access to network resources needs to be addressed at the international level, since traditional geographic boundaries have no place on the Internet.

References

Barron, Billy <BILLY@VAXB.acs.unt.edu>. (1991 and ongoing). UNT's accessing online bibliographic databases. Denton, TX: University of North Texas. Available via FTP from FTP.UNT.EDU in the /library subdirectory. Retrieve the file LIBRARIES.TX2 for pure ASCII version.

Barron, Billy <BILLY@VAXB.acs.unt.edu>. (1992). Telnet to: quake.think.com and login with wais. Find the entry for the "hytelnet" server, hit the space bar to mark, select keyword to search, and hit enter. The file(s) containing your keyword will be retrieved.

Duggan, Richard <DUGGAN@BRAHMS.udel.edu>. (1990). CATALIST. Newark, DE: University of Delaware. Available via FTP from FTP.UNT.EDU in the library subdirectory.

Fogel, Earl <access.usask.ca>. (1992). Direct all enquiries fogel@skyfox.usask.ca.

Hu, Tsung (1992). TELEMATE. Available via FTP from WUARCHIVE.WUSTL.EDU in the mirrors/msdos/modem subdirectory in three files: tm301-1.zip, tm301-2.zip, tm301-3.zip. Change file type to "binary" when fetching.

Internet Gopher. (1992). Telnet to: consultant.micro.umn.edu and login with gopher. Emulate vt100. Use menus to find "libraries" entry.

Larson, Neil. (1989). HyperRez. California: MaxThink. Available via FTP from WUARCHIVE.WUSTL.EDU in the mirrors/msdos/hypertext subdirectory as hyperrez.arc. Change file type to "binary" when fetching.

LIB_HYTELNET. (1990). To subscribe, send an e-mail message to scott@sklib.usask.ca, with the subject "Add me to LIB_HYTELNET"

QEdit advanced v2.15. (1991). Marietta, GA: SemWare. Available via FTP from WUARCHIVE.WUSTL.EDU in the mirrors/msdos/qedit subdirectory as qedit215.zip. Change file type to "binary" when fetching.

Scott, Peter <SCOTT@SKLIB.usask.ca>. (1990). HYPERVAX. Saskatoon, SASK: Available via FTP from WUARCHIVE.WUSTL.EDU in the mirrors/msdos/hypertext subdirectory as hyprvx.zip. Change file type to "binary" when fetching.

Scott, Peter <SCOTT@SKLIB.usask.ca>. (1991a). HYDOS. Saskatoon, SASK: Available via FTP from WUARCHIVE.WUSTL.EDU in the mirrors/msdos/hypertext subdirectory as hydos10.zip. Change file type to "binary" when fetching.

Scott, Peter <SCOTT@SKLIB.usask.ca>. (1991b). HYENVOY. Saskatoon, SASK: Available via FTP from WUARCHIVE.WUSTL.EDU in the mirrors/msdos/hypertext subdirectory as hyenvoy1.zip. Change file type to "binary" when fetching.

Scott, Peter <SCOTT@SKLIB.usask.ca>. (1992). HYTELNET. Saskatoon, SASK: Available via FTP from WUARCHIVE.WUSTL.EDU in the mirrors/msdos/hypertext subdirectory as hyteln50.zip. Change file type to "binary" when fetching.

St. George, Art <STGEORGE@UNMB.bitnet>. (1990 and ongoing). Internet-accessible library catalogs and databases. Albuquerque, NM: University of New Mexico. Available via e-mail from LISTSERV@UNMVM.BITNET. Send the message: GET LIBRARY PACKAGE.

Appendix*

➢ To Retrieve HYTELNET from the University of Saskatchewan:

At your system prompt, enter:ftp access.usask.ca When you receive the Name prompt, enter: anonymous When you receive the password prompt, enter your Internet address. When you are at the ftp> prompt, enter: binary At the next ftp> prompt, enter: cd hytelnet/pc Then enter: get hyteln50.zip

After the transfer has occurred, either proceed with the instructions below to retrieve the UNZIP utility (which you need unless you already have it) or enter: quit

The Hytelnet program is archived using a ZIP utility. To unarchive it, you must be able to "unzip" the file. If you have the file PKUNZIP.EXE, it will unarchive the HYTELN50.ZIP file (see below for instructions). If you do not have it, you may retrieve it by following these instructions:

➢ To retrieve PKUNZIP.EXE:

Use the above instructions for connecting to access.usask.ca At the ftp> prompt, enter: binary Then enter: cd hytelnet/pc Then enter: get pkunzip.exe After the transfer has occurred, enter: quit

- *To download it to your PC:*

Because of the plethora of PC communications programs, I will not attempt to give step-by-step instructions here. You should check the instructions for your software for downloading a binary file from your Internet account to your PC.

➢ To unarchive HYTELN50.ZIP:

Make a new directory on your hard disk (e.g., mkdir hytelnet) Copy PKUNZIP.EXE and HYTELN50.ZIP into the new directory Make sure you are in that directory, then enter: pkunzip HYTELN50 It will then unarchive HYTELN50.ZIP, which contains the following files: HYTELNET.ZIP and READNOW.!!!

The file READNOW.!!! gives full instructions for un-archiving HYTELNET.ZIP. Simply put, you MUST unZIP the file with the -d parameter so that all the subdirectories will be recursed.

➢ Loading HYTELNET:

At the DOS prompt (in the HYTELNET parent directory), type HR to install the program in memory. After it loads, hold the Ctrl key down and depress the Backspace (<-) key.

It is a memory-resident program that should be invoked before you load your communications program. Have it sit in the background until you need to find a Telnet address. To invoke the program just hit the Control and Backspace keys; then follow the directions. When you have read the site information, either hit the Escape key to return the program to the background, or hit Alt-T to remove it from memory.

Program size: 16065 bytes (HyperRez on disk). ASCII file size: Maximum size is 20K (set by text buffer). Maximum recall: Remembers. Right-arrow jumps 64 levels deep.

Essential files for running the program:

Program: HR.EXE (HyperRez program) Select hot-key: HRK.EXE Title ASCII file: START.TXT HyperRez F1 file: HELP.TXT Instructions: READ.ME

*Editor's Note: The following section deletes periods at the end of some sentences so as not to confuse punctuation with command syntax.

Resource Discovery in an Internet Environment—the Archie Approach

Peter Deutsch

New resources and services are being added to the network daily. The number of prospective users of these resources is expanding rapidly, but problems arise when individuals attempt to identify, locate, and access networked information in today's dynamic environment. This paper describes Archie, an electronic indexing service for locating information that exists on the Internet. The author describes the Archie service in the context of the Resource Discovery Problem and discusses enhancements that are planned for Archie.

The development of the Internet represents perhaps one of the greatest collaborative research efforts ever undertaken. Originally created to connect a relatively small group of like-minded researchers in computer science and engineering, the Internet has now grown to support a closely coupled community of millions of users having access to thousands of networks and hundreds of thousands of machines located throughout the world.

Much of the early effort in the design and deployment of the Internet by necessity concentrated on such low-level issues as development of needed protocols and hardware, with little time or energy left for the more abstract problems of providing specific user-level services in the new distributed computing environment. Thus, for much of the first ten years, the Internet was used for little more than remote login, electronic mail and file transfer from remote archives.

This situation is now changing. The repertoire of network services now spans a wide range, from the Usenet news bulletin board service to on-line library catalogues, campuswide information systems and even on-line weather information services. It may be argued that, as the network has grown from a collection of hundreds of machines to one of hundreds of thousands of machines, a fundamental shift in focus has occurred among users. Rather than seeing themselves as interacting primarily with other individuals on the net, users more and more have come to see themselves as interacting with "the net" itself, with the vast pool of machines and their associated resources functioning as a virtual provider of electronic goods and services.

As the emphasis on Internet development shifts from implementation of basic connectivity to the provision of a range of network services, new problems and new approaches to solving those problems arise. The "Archie group," a collection of volunteers based at McGill University in Montreal, has developed "Archie," an automated service that fetches and indexes information distributed across the Internet, making it available to Internet users in a variety of ways. In this article I examine the Archie service, along with some of the general issues and problems involved in providing such services in the current Internet environment.

***Peter Deutsch** <peter@cc.mcgill.ca> Computing Centre, McGill University, Room 200, Burnside Hall, 805 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6, is one of the architects of the Archie system and also President of Bunyip Information Systems, a company formed by the creators of archie to develop follow-on versions of the service. He is also co-chair (with Archie co-designer Alan Emtage) of the Internet Anonymous FTP Archives Working Group of the Internet Engineering Task Force and one of the organizers of "the Archie group," a collection of volunteers working to expand and improve the Archie project.*

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 45–51. © 1992 Meckler

The Resource Discovery Problem

As the number of users and hosts continue to grow, both individual users and potential service providers have come to recognize that a major challenge exists in identifying the existence and location of services and service providers in a distributed environment of hundreds of thousands of machines. This problem, the so-called Resource Discovery Problem, must be adequately addressed if we are to move towards a true Internet-wide model of resource delivery. The problem of locating the specific hosts that offer a needed on-line library catalogue service, technical report, or piece of software is compounded by the fact that the current Internet is growing at a phenomenal rate. The most recent count of connected machines (estimated by counting registrations in the Domain Name System using automated software) puts the current number of machines at over 700,000 (up from some 376,000 this time last year) with growth now running at 30 percent every three months (Lottor, 1992).[1]

Locating a specific service provider in this vast and growing sea of hosts has become a serious problem, one that must be addressed if our users are to become comfortable in viewing the Internet in terms of the services it can provide and not simply as the hardware needed to make it work. Several researchers have attempted to model the problem of locating and accessing information. Yeong (1991), addressing of the problems of networked information retrieval, speaks of "Discovery, Searching and Delivery," while Schwartz (1991), defines the problem in terms of "Class Discovery, Instance Location and Access." Other researchers are working to provide methods for easing the burden of information management while also providing accessing methods (Kahle, 1989; Neumen, 1989). I will use Schwartz's terminology for discussion of the Resource Discovery Problem in this article, although I will sometimes depart from his model in certain respects.

The act of Class Discovery refers to seeking out a specific type of service in a larger community of service providers. Thus, a user might wish to locate "anonymous FTP archive sites" (that is, hosts that provide universal access to their collections of information). These sites offer a wide range of information, including technical reports and other publications, software, and data, using the convention of a special "anonymous" user code that requires no password while permitting access to the FTP file transfer protocol.

Once the existence of a specific class of service has been established, a user can proceed to Instance Location. Continuing with the above example, once the location of various anonymous FTP archive sites for storing software has been established, a user could then search for an accessible copy of the password protection program "npasswd." This particular program exists at a number of locations throughout the Internet, and presumably the user would be interested in locating a version that could be copied to his or her site, perhaps taking into consideration such factors as the need to minimize transmission time or network load.

Until recently, such a search would be a time-consuming and difficult challenge, for there was no universal registry of archive sites and no method for searching short of logging on and examining the filenames at each of hundreds of sites in turn. The particular problem of indexing anonymous FTP archives was addressed successfully by the Archie project, developed by students and volunteers at McGill University in Montreal. The architecture of the current Archie system is described in Emtage and Deutsch (1992).

Originally created to track the contents of anonymous FTP archive sites, the Archie service is currently being expanded to include a variety of other on-line resource listings.

In this example of anonymous FTP archives, access would then be achieved using the ftp protocol to transfer the file from the appropriate archive to the user's site.

One Method for Instance Location: The "Archie" Service

The Archie service is a collection of tools that, taken together, provide an electronic indexing service for locating information in the Internet environment. One identifying feature of Archie is that its indexing information is actually gathered directly from

primary sources on the net by automated tools, with this information being periodically updated in a proactive manner. This assures users that indexing information is reasonably current and accurate at all times. Originally created to track the contents of anonymous FTP archive sites, the Archie service is currently being expanded to include a variety of other on-line resource listings. The basic Archie model is simple and flexible, making it suitable for tracking any periodically changing collection of information distributed across the Internet, provided that this information is accessible on the net to Archie's automatic data gathering component.

The Archie system offers a simple client-server model of Internet Instance Location. The server automatically gathers the information on a regular basis, and users contact the server using any one of several client programs to perform searches on this information when needed.

Class Discovery to locate an Archie server is not actually addressed (it is assumed users are aware of Archie's existence) but once an Archie server is located the task of locating specific archive sites is handled by the server. This allows the user to perform Class discovery on archive sites and quickly locate specific instances of information through Archie index searches. In addition, the use of proactive data gathering allows users to place a great deal of confidence in this secondary source of information, since its information is derived automatically from primary sources on a regular basis.

Accessing Archie

Several methods exist for accessing the Archie databases. Interactive sessions can be initiated using the basic telnet command to an Archie server, although more efficient access is available through any one of several client programs now available for Xwindows, NeXTStep, DOS, or VMS environments. Alternatively, the Archie files database can be accessed directly through the Prospero distributed file system. Finally, users can send queries through electronic mail, provided they can at least gateway electronic mail messages onto the Internet.

The existence of the Archie service allows users of anonymous FTP to limit their Instance Location searches to a set of questions which are directed to one of a small number of known Archie servers; these in turn offer pointers to specific Internet service providers. Once the existence and location of specific Instance information have been determined using Archie, the existing FTP protocol can be used for final access. Many of the Archie GUI-based clients have integrated FTP support directly into their programs, creating a generalized archive indexing and access tool.

Trying out the Archie service

Users with direct Internet connectivity can try out an interactive Archie server using the basic Telnet command (available at most sites). To use, telnet to the host "archie.mcgill.ca" [132.206.2.3] and login as user "archie" (there is no password needed). A banner message giving the latest developments and information on the Archie project will be displayed, and then the command prompt will appear. First-time users should try the "help" command to get started. The "servers" command will list all active Archie servers, so you can pick one closer to your site for improved response time.

Users with only e-mail connectivity to the Internet should send a message to "archie@archie.mcgill.ca," with the single word "help" in either the subject or body of the message. You should receive back an e-mail message explaining how to use the e-mail Archie server, along with details of an e-mail-based ftp server that will perform the actual FTP transfers for you.

Additional Archie client programs may be obtained through anonymous FTP and are stored on archie.mcgill.ca in the subdirectory "archie/clients." Currently, there are stand-alone Telnet and Prospero clients, as well as Perl, Xwindows, and NeXTstep clients that use the Prospero server protocol but offer a more "user friendly" front end. Clients for MS-DOS and VMS are also available.

Documentation for the Archie system is stored on the archie.mcgill.ca archive under the directory "archie/doc." This includes copies of the Archie manual, a brief description of the service, an architectural overview, and this article.

User feedback on all aspects of the Archie project is welcome. The Archie project is still entirely a volunteer affair, and user feedback is invaluable in helping us to focus our limited resources where they will do the most good. Feel free to send comments and suggestions to one of the addresses at the end of this article.[2]

The Archie Service Today

Currently, Archie tracks the contents of over 900 anonymous FTP archive sites containing over 1.6 million files throughout the Internet. Collectively, these files represent well over 105 Gigabytes (92 billion bytes) of information, with additional information being added daily. Anonymous FTP archive sites offer software, data, and other information that can be copied and used without charge by anyone with connection to the Internet.

As mentioned, the Archie server automatically updates the listing information from each site on a regular basis. The frequency of updates is easily configurable, and the various Archie site administrators have experimented with a variety of updating schemes. The most common scheme is a round-robin arrangement where each site is updated over a period of weeks. Some Archie sites actually perform this updating on a daily basis for certain key sites. Work is also underway to ensure synchronization of the multiple Archie databases. There are now nine Archie servers in operation around the world, with more on the way.

A growing range of additional service providers is anticipated, including dedicated information servers for each host that shares the architecture's object-oriented approach, thus allowing them to be integrated with the Class Discovery and Indexing layers.

In addition to the anonymous FTP files database, Archie offers the "whatis" descriptions database. This database provides the name and a brief synopsis for over 3,500 public domain software packages, datasets, and informational documents located on the Internet. This database is not yet generated or maintained automatically, but this is planned for the next release of the system.

Additional Archie databases are also scheduled to be added in the next release. Planned offerings include indexed collections of abstracts and software descriptions that will be automatically updated in the same manner as the filenames database. Additional listings will summarize the names and locations of on-line library catalogue programs, publicly accessible electronic mailing lists and archive sites for the most popular Usenet "newsgroups" or bulletin boards.

Work is currently underway in the Internet Anonymous FTP Archives Working Group (IAFA-WG)[3] of the Internet Engineering Task Force to develop a standard encoding method for information to be collected by services such as Archie. Once a general mechanism for encoding such information is completed, the way is clear for the Archie system to become a generalized information indexing system for the entire Internet. Suggestions for additional descriptions or locations databases are also welcomed and should be sent to the Archie developers at one of the addresses at the end of this paper.

Future Work

The primary goal for the next release of Archie is to extend the range of information tracked, but a number of other improvements are also planned. These include additional access methods, additional search methods to speed interactive searches, and extensions to the basic telnet interface to support these methods.

The current Archie system allows access to a limited number of databases. As additional databases are added, plans call for more database access methods. Current plans call for a Wide Area Information Servers (WAIS) interface (the WAIS system is discussed below) within Archie to provide rapid indexing and access to large textual databases in the next release. Additional interfaces are expected in future releases.

The current version of Archie offers a variety of search methods, including exact match, case-insensitive searching, regular expression searching, and more. All return a large collection of information on each file that matches the specified search pattern, and a poorly chosen search term can generate a large amount of extraneous information.

The next release of Archie will add another "fast match" search option that will return only the appropriate filename matches, without the other associated information. It is anticipated that the user will scan these matches manually and select only

those that seem to actually refer to the desired information. If the entire collection of information on these files is still desired, they can be fetched using an efficient exact match lookup, speeding search time while lowering the load on the Archie servers.

Work is also underway to reimplement the Archie telnet client to use a client-server architecture internally. This is needed to allow this interface to access the planned additional databases and access methods.

All the above changes are planned for the next release of Archie. Further enhancements for later versions include the implementation of an "update interrupt" and callback mechanisms. The first mechanism will allow archive sites to trigger a site update automatically whenever changes are made to their collections. This will eliminate the current latency that exists between when a site is updated and when the site is next visited by the Archie updating tools. The second change is intended to allow users to register requests for specific items with an Archie server. Whenever a site update is performed, any search term that is affected by the update will generate a message to the user, allowing immediate notification of the arrival or departure of specific information from the Internet.

Beyond Archie

The Archie service is one component of a larger information publication architecture currently under development by the Archie group. This architecture is intended to address all three components of the Resource Discovery Problem, providing a generalized resource discovery and an access mechanism for the Internet.

This system models information as collections of typed objects, with a specified collection of attributes for each type of information available. The basic architecture of the system consists of the Resource Information Service (RIS), the Resource Indexing Service, and dedicated Information Providers.

The Resource Information Service layer (RIS) provides the needed mechanism for Class Discovery. It is intended that this service act as a registry of services available on the net. It is anticipated that this registry would include a brief text abstract of the service that could itself be indexed and searched, as well as attributes that would provide such information as service type, location, and other access information. Users (or automated user tools) would be able to query the RIS for information about specific types of services.

The Indexing Services layer corresponds to an extension of the existing Archie service and thus provides a mechanism for Instance location. It is anticipated that in the future we will see a trend toward specialized indexing services in response to scaling and performance concerns. By dedicating Archie-like servers to specific portions of the information space, we avoid potential bottlenecks while also limiting the search for specific types of information, thus improving performance.

A comparison can be drawn between such indexing services and the role of magazine editors in the existing publishing industry. The editor acts as a filter, selecting a specific type of information for inclusion in a specific publication. Users are spared the necessity of wading through inappropriate submissions (for example, car reviews in a home furnishing magazine) while they are granted access to a timely collection of useful information on the subject of their choice (or in this case, pointers to information, as what is served is the indexed directories of specific service providers).

The final layer in this model is the Service Providers themselves, which allow users to actually access the desired information. Current service providers include anonymous FTP archives, news servers, WAIS source servers, and others. A growing range of additional service providers is anticipated, including dedicated information servers for each host that shares the architecture's object-oriented approach, thus allowing them to be integrated with the Class Discovery and Indexing layers.

It is anticipated that intelligent user agents will be developed that are aware of and make use of this architecture, thus providing tools that can perform the three steps of Resource Discovery automatically.

Other Tools for Resource Discovery and Access

A number of different information discovery and delivery tools have been developed, with most such tools offering facilities for Instance Discovery, Access, or Information Management. Some have been with us from the early days of the Internet, whereas others have appeared only in the past couple of years.

The Domain Name System (DNS) (Mockapetris, 1987) was an early example of a network-wide distributed database system. Primarily designed to

perform translation from fully qualified domain names to IP addresses, it is also used to distribute information about host hardware, operating systems configurations, and electronic mail exchanger addresses. DNS has been an operational success, having expanded continuously since its inception to now cover over 700,000 machine names. Despite this success, there are problems.

Maintenance of the system is distributed, with the required information entered into flat text files (usually by hand) at the site of each authoritative subdomain server. This can lead to inconsistencies and errors in the database that can only be corrected through human intervention. There is no internal consistency checking of this information by the system itself (for example, to verify that registered hosts actually exist on the net). Another problem can arise during operation. If the authoritative server for a particular subdomain becomes unreachable, then users will find that they cannot perform hostname to address conversion. In this case, users can find themselves unable to access a host, even though that particular host is available.

This problem can be alleviated by the use of suitably chosen replicating servers (or by using the IP address itself, where it is known), but the configuration and operation of these replicated servers are not automatic and are again prone to human error.

Despite these drawbacks, DNS illustrates the feasibility of large network-based database server applications in an Internet environment. Distributed file systems such as "NFS: Network File System Protocol Specification" (1989) and Prospero allow site administrators to distribute file systems across multiple hosts in a network environment.

Among other features, the Prospero file system (actually one component of the larger Prospero virtual computing environment now under development) provides the capability for creating customized views of available files through user-specified links. This configuration information is itself a form of value added processing of the file system information over and above the contents of the individual files themselves. Such a customized view can then, in turn, be exported and accessed by others, aiding in both the Instance Location and Information Management areas.

Internet white pages directory services (Sollins, 1989), are intended to provide the on-line equivalent of a white pages phone book. Such services aim at providing users with access to user login names, e-mail addresses, and other contact information. A White Pages Directory Service project based on the X.500 protocol is described in by Deutsch (1988).

Work on the X.500 project is carried out through a number of forums, including the Internet Engineering Task Force, ISO standards committees, and the U.S. government GOSIP program.

The Wide Area Information Servers (WAIS) system is an example of a network-based document indexing system that has proved useful for accessing large collections of textual data. The WAIS system, based on the ANSI Z39.50 protocol standard, provides an indexing and search mechanism that allows the user to rapidly perform keyword searches on documents that can be tens or hundreds of megabytes in size. The WAIS system can locate the desired keywords and then return the appropriate portion of the document to the user's machine, addressing both the instance location and information mangement portions of the problem.

Conclusion

The year 1991 witnessed an explosion of interest in using the Internet to develop and deliver user services. This author believes that projects such as Archie, though limited in scope, offer a significant glimpse of the power that Internet connectivity will bring to future computer users. The promise is of a range of imaginative services that fully exploit this new environment while offering high quality and near universal accessibility.

Significant problems still remain for those who would develop and deploy such services, not the least of which is the nagging question of how to fund services in an environment where direct usage charges are still not accepted by most would-be consumers and where such services, if offered for a fee, would still violate "appropriate use" policies of many of the connecting networks.

The current model calls for sites to offer such services on a volunteer basis, with each site supposedly offering back services in some measure that is proportional to what the net has brought to their site. While this method has worked to date for such services as anonymous FTP, there is some doubt that this can continue to work as the potential user base grows, threatening to quickly choke off any service in its own success.

From our own experience with Archie, a huge part of the work in establishing such services is not technical, but political in nature. In our case, we had to persuade a number of sites to donate equipment and personnel to run servers, persuade network connectivity providers to accept the huge increase in traffic that resulted, and persuade our own institution to allow us to continue working on what was supposed to be a hobby project of limited scope. To a degree we were successful, but the process is not one for the faint of heart, nor am I sure that it is one that is ultimately fair to either service providers or users.

We need to develop additional mechanisms to pay for Internet user services. With the continued growth of the Internet, we can expect to see, perhaps, funding from regional service providers, direct charging mechanisms to users or even, sadly, nonuniversal access. Such may be the way we want the Internet to evolve, but it is hoped that whatever result we arrive at, it is the product of reflection and debate, and not the result of chance and circumstances.

I am optimistic that we will overcome such nontechnical problems just as we have overcome the technical ones. The next ten years promise, in their own way, to be as spectacular as the past ten.

Notes

1. A number of the references in this article are "RFCs," or Request For Comments. These are documents that have gone through the Internet Engineering Task Force review process and have been accepted as part of the body of standards on which the Internet is built. RFCs are themselves available from a number of sites on the Internet, including:

- ftp.nisc.sri.com
- nis.nsf.net
- nisc.jvnc.net
- venera.isi.edu
- wuarchive.wustl.edu
- nic.ddn.mil

Instructions for retrieving RFCs may be found in the file "in-notes/rfc-retrieval.txt" on VENERA.ISI.EDU.

2. The email address "archie-group@archie.mcgill.ca" reaches the archie implementation team, while the mailing list "archie-people@archie.mcgill.ca" has been established for those wishing to keep informed of developments on the archie project. Send your subscription requests to "archie-people-request@archie.mcgill.ca"

3. One can join the IAFA-WG mailing list by sending mail to <iafa-request@cc.mcgill.ca>.

References

Deutsch, D. (1988, June). *An introduction to the X.500 series network directory service.* Cambridge, MA: BBN Laboratories.

Emtage, A., & Deutsch, D. (1992). Archie—An electronic directory service for the Internet. *In Proceedings of the Winter 1992 USENIX Tech Conference* (pp. 93-110). Berkeley, CA: USENIX Association [2560 Ninth St., Suite 215, Berkeley, CA 94710].

Kahle, Brewster (1989, November). *Wide area information servers concepts* (Thinking Machines technical report TMC-202). (Available via anonymous ftp: /pub/wais/doc/wais-concepts.txt@quake.think.com. OR wais server wais-docs.src.)

Lottor, M. (1992, January). *Internet growth* (1981-1991) (Internet RFC 1296).

Mockapetris, P. (1987, November). *Domain names—concepts and facilities* (Internet RFC 1034).

Neuman, C. (1989). The virtual system model for large distributed operating systems. Seattle: University of Washington, unpublished PhD dissertation.

NFS: Network file system protocol specification. (1989, March). (Internet RFC 1094).

Schwartz, Michael. (1991, November). *Resource discovery in the global Internet.* (CU-CS-555-91). Bolder, CO: University of Colorado.

Sollins, K. (1989, June). *A plan for Internet directory services* (white pages). (Internet RFC 1107).

Yeong, W. (1991). *Towards networked information retrieval* (Tech. Report 91-06-25-01). Reston, VA: Performance Systems International, Inc.

World-Wide Web: The Information Universe

Tim Berners-Lee, Robert Cailliau, Jean-François Groff, and Bernd Pollermann

The World-Wide Web (W^3) initiative is a practical project designed to bring a global information universe into existence using available technology. This article describes the aims, data model, and protocols needed to implement the "web" and compares them with various contemporary systems.

The Dream

Pick up your pen, mouse, or favorite pointing device and press it on a reference in this document—perhaps to the author's name, or organization, or some related work. Suppose you are then directly presented with the background material—other papers, the author's coordinates, the organization's address, and its entire telephone directory. Suppose each of these documents has the same property of being linked to other original documents all over the world. You would have at your fingertips all you need to know about electronic publishing, high-energy physics, or for that matter, Asian culture. If you are reading this article on paper, you can only dream, but read on.

Since Vannevar Bush's article (1945), men have dreamed of extending their intellect by making their collective knowledge available to each individual by using machines. Computers give us two practical techniques for human-knowledge interface. One is hypertext, in which links between pieces of text (or other media) mimic human association of ideas. The other is text retrieval, which allows associations to be deduced from the content of text. In the first case, the reader's operation is typically to click with a mouse (or type in a reference number). In the second case, it is to supply some words representing that which he desires. The W^3 ideal world allows both operations and provides access from any browsing platform.

Reality

Existing research projects and commercial products are not far form achieving parts of this dream. The Xanadu system is an ambitious distributed hypertext project. Existing hypertext systems (see for example *Beyond Hypertext*, 1990, Kahn, et al., 1988, & Nelson, 1988) tend to be restricted to the local or distributed file system and they often are developed with a limited set of platforms in mind. Contemporary information retrieval and access systems such

Tim Berners-Lee *<trimbl@info.cern.ch> has a background in text processing, real-time software, and communications. He developed his first hypertext system, "Enquire," in 1980 for his own use and he started the World-Wide Web initiative in 1989.*

Robert Cailliau *<cailliau@cernvm.cern.ch> has been interested in document production since 1975 and is a long-time user of Hypercard. He ran CERN's Office computing Systems group from 1987 to 1989 and joined Tim Berners-Lee in 1990 to start the World-Wide Web.*

Jean-François Groff *<jfg@info.cern.ch> graduated from the French National Institute of telecommunications and came to CERN after working for Intelsat. As a strong believer in the future of networked hypertext, he joined the World-Wide Web team in 1991.*

Bernd Pollermann *<bernd@cernvm.cern.ch> has experience in managing large bases of information and is responsible for the index of CERN's computer center documentation, now included in the Web. He is also the editor of CERN's Computer Newsletter.*

The mail address for all authors is: CERN, 1211 Geneva 23, Switzerland, Phone: +41 22 767 3755, or +41 22 767 7155 for fax.

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 52–58.

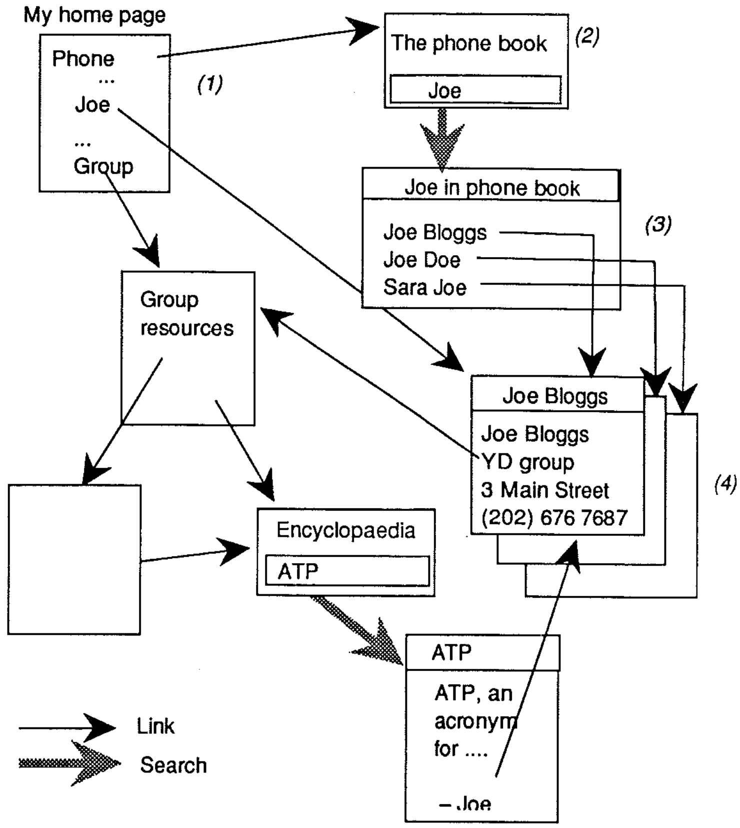
as Alex (Cate, 1992), Gopher (Alberti, et al. 1991), Prospero (Neuman, 1992), and WAIS (Davis, et al., 1990) cover a wide area without the hypertext functionality. Merging the techniques of hypertext, information retrieval, and wide area networking produces the W³ model.

The W³ Data Model

The W³ model uses both paradigms of hypertext link and text search in a complementary fashion, for neither can replace the functionality of the other. Figure 1 shows how a personalized web of information is built from these operators.

Features to note are:

- Information need only be represented once, as a reference may be made instead of making a copy.
- Links allow the topology of the information to evolve, so modeling the state of human knowledge at any time is without constraint.
- The web stretches seamlessly from small personal notes on the local workstation to large databases on other continents.
- Indexes are documents, and so may themselves be found by searches and/or following links. An index is represented to the user by a "cover page" that describes the data indexed and the properties of the search engine.
- The documents in the web do not have to exist as files; they can be "virtual" documents generated by a server in response to a query or document name. They can therefore represent views of databases, or snapshots of changing data (such as the weather forecasts, financial information, etc.).



The W³ model involves hypertext links and index searches. The reader starts at the home page (1) and quickly uses his own links, group-wide or public links, to find resources. Indexes such as the phone book (2) are represented as documents with the possibility of inputting search words. The result is a virtual hypertext document (3) which points to the documents found (4).

Figure 1. A web of links and indexes

A pleasing and useful aspect is that almost all existing information systems can be represented in terms of the W³ model. A menu becomes a page of hypertext, with each element linked to a different destination. The same is true of a directory, whether part of a hierarchical or cross-linked system. The notion of many named indexes within the web allows a given search engine and database to be visible with several different addresses, each representing different options for the search algorithm. For example, the index `/library/books/ti+au/substring` may give a title and author search, whereas

`/library/books/text/exact` may give an exact-word full-text search. Addresses are discussed in more detail below.

Publishing

From the information provider's point of view, existing information systems may be "published" as part of the web simply by giving access to the data through a small server program. The data itself, and the software and human procedures that manage it, are left entirely in place. This approach has allowed, for example, a mainframe-based document storage and index system to be opened up to all platforms in the organization. To see how this is done requires a brief overview of the W^3 architecture.

W^3 Architecture

Hypertext and text retrieval systems have been available for many years, and a valid question is why a global system has not already come into existence. Traditional answers to this question are the lack of:

- a common naming scheme for documents
- common network access protocols
- common data formats for hypertext

Most research in hypertext systems (the Xanadu project excepted) have focused on the user interface and authoring questions rather than on the questions of wide-area and long-term distribution. These architectures have assumed that users share a common application program running on computers (often of the same type) that share a common file system. However, the W^3 architecture must cope with a widely distributed heterogeneous set of computers running different applications that use different preferred data formats. This requires a client-server model. The client has the responsibility for resolving a document address into a document using its repertoire of network protocols. The server provides data in a simple hypertext or plain text form, or, by negotiation with the client, in any other data format.

It may be more difficult initially to develop a generic hypertext browser than a specific front-end for a particular information system. However, the decoupling of the client and server programs by the "information bus" pays off as more clients and servers are plugged in and universal readership is achieved. Writing a server for new data is generally a simple task because it requires no human interface programming.

Document Naming

The fulcrum on which the document universe rests is the scheme for naming documents. A document name provides a method for the client to find the server and for the server to find the document. In the W^3 model, a name can also specify a part of the document to be selected from the displaying application.

Although a document name is normally hidden in the hypertext syntax transferred over the link, in practice it must sometimes be referred to by people, and passed through applications (such as mail) that are not yet hypertext-aware. Therefore, ideally it must be composed of printable characters and manageably short.

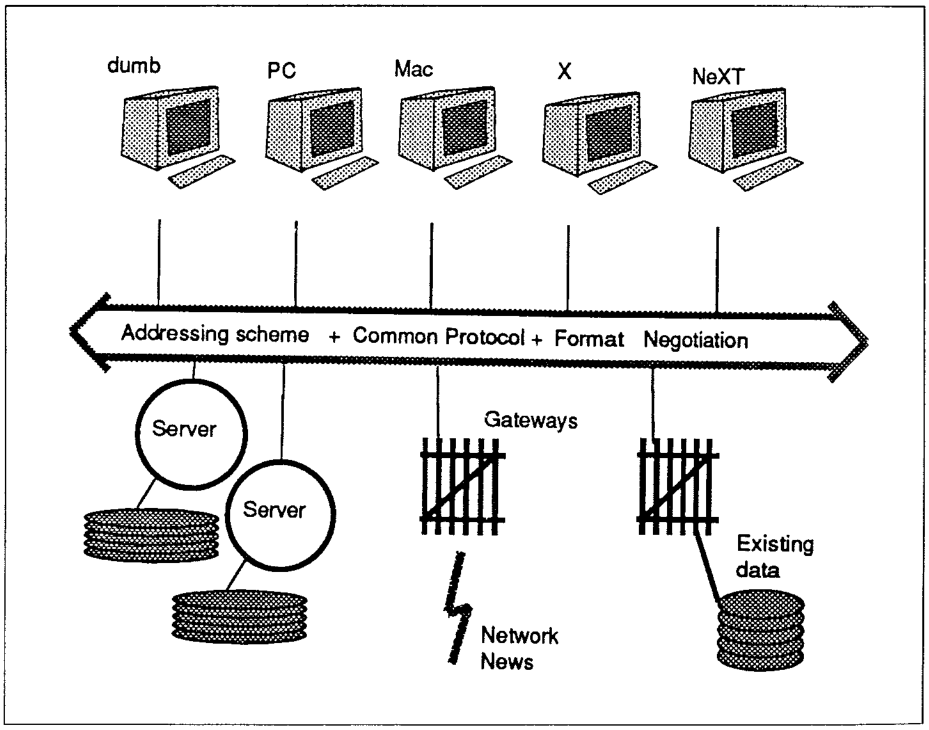


Figure 2. The W^3 architecture in outline

Any lasting reference to a document must be a logical name rather than a physical address. That is, it should refer to a document's registration and some "publishing" organization rather than any physical location, so that its location may later be moved. The client is therefore prepared to follow several stages of translation by name servers before finding a final document server. Similarly, a document name should not contain any information that is transitory, such as the particular formats available for a document or its length.

The W³ naming scheme fulfills these requirements but is otherwise open to the addition of new protocols as technology evolves. For this purpose a prefix is used to identify the protocol (and therefore naming scheme) to be used. Clients who do not have that protocol in their repertoire refer to a gateway for translation.

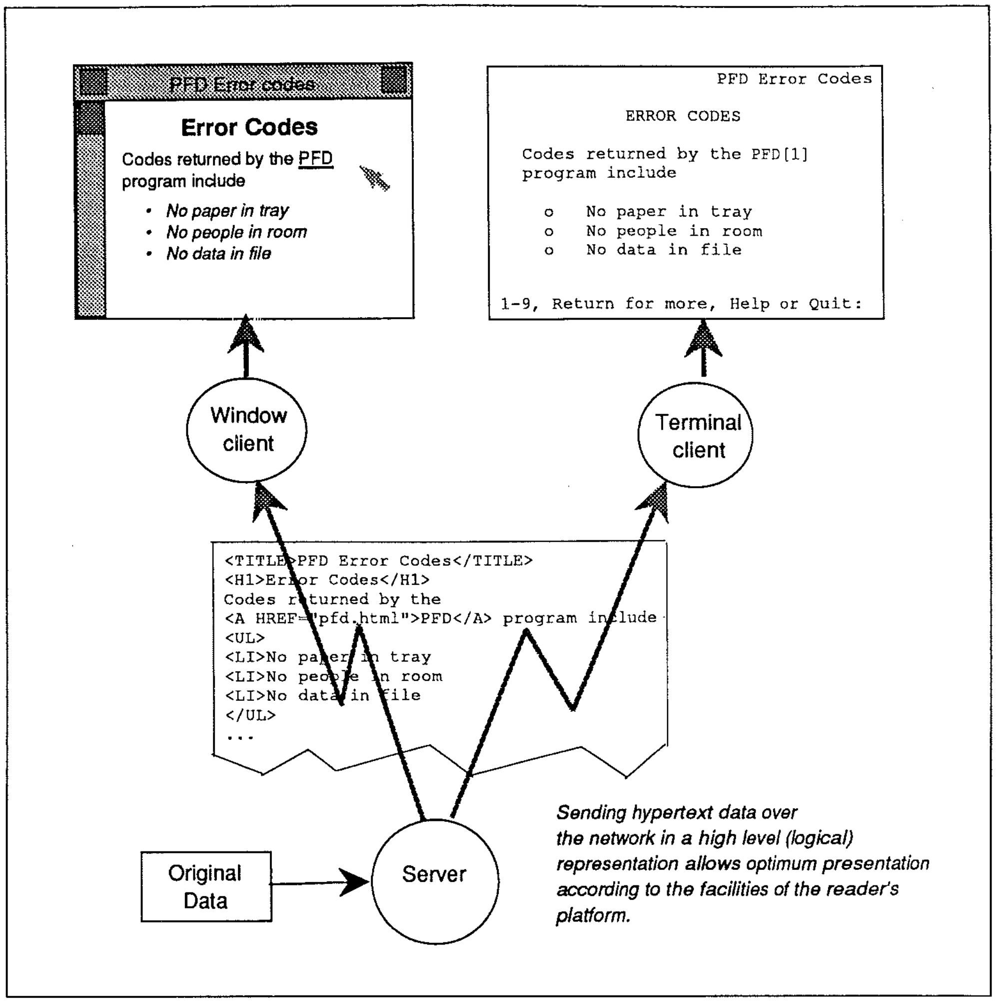


Figure 3: A schematic illustration of the encoding of hypertext data. The link is represented in the window by underlining, on the terminal by a reference number.

Protocols

The W³ clients are built on a common core of networking code for information access. This core provides access using widely deployed internet protocols such as:

- File Transfer Protocol—FTP (Postel & Reynolds, 1985)
- Network News Transfer Protocol—NNTP (Kantor & Lapsley, 1986)
- Access to mounted file systems

A new search and retrieve (SR) protocol, known as HTTP, was found to be necessary. Faster than FTP for document retrieval, HTTP also allows index search. HTTP is similar in implementation to the Internet protocols above and similar in functionality to the WAIS protocol. Some differences are discussed below.

Document Formats

The Dexter data model of hypertext (Halasz & Schwartz, 1990) provides a conceptual model for hypertext systems and the HyTime standard (Goldfarb, 1991) formalizes hypertext at a high level. The W³ project defines a concrete syntax in the SGML style for basic hypertext as it is used for menus, search results, and online hypertext documentation.

Even W^3 browsing application is able to parse this simple format (see Figure 3). In the pilot phase of the project, this format was all that was required, but in the second phase, format negotiation between client and server will allow the exchange of information in any medium using any mutually acceptable representation.

WAIS and the Web

From the point of view of the W^3 dream, the WAIS protocol represents a significant advance on the search and retrieve protocol standard Z39.50/ISO-10163 by being stateless and introducing a persistent name. The document names used are local to the containing database, but these names may be appended to the database name and host address to form a universal W^3 address. In this way, WAIS indexes and servers can be represented in the web. A gateway program, running at CERN and available for general use, provides this mapping. The WAIS model also uses separate "source" files to describe indexes. The WAIS-W^3 gateway keeps caches of these files, using them to build descriptive "cover pages" for indexes.

The current WAIS model requires that the results of a search point to documents available from the same server. That is, the same server is responsible for indexing and actually providing the data. In the W^3 world this restriction does not exist. A practical advantage of this approach is that, as Yeong (1991) points out, a large multimedia document may be most efficiently retrieved from a different host and by using a different protocol from that used for the original query. Furthermore, as online information proliferates, an important function is that of "third party" reviewers, indexers, and overview writers who refer to data they do not actually hold. It is expected that these services will be a key to the control of the information explosion and a valuable asset to the community.

A W^3 user builds a personalized web of information by making links from his own notebook into the web. He can make a link to the result of a search, so that the next time he follows the link the search is re-evaluated. This is the equivalent of storing a WAIS "question"—there is a good mapping between the models. The W^3 clients do not currently support relevance feedback, although it is not alien to the model.

There are two occasions when hypertext would particularly enhance the WAIS model. First, users often would like to be able to browse through available WAIS indexes. Both WAIS and W^3 regard indexes as documents and therefore allow them to be found using the same techniques as for documents. In fact, the WAIS-W^3 gateway allows a W^3 hypertext overview to be made with pointers to WAIS indexes. Second, when one has found a piece of text, WAIS delivers just that part of file that has been found. Very often one would like links to surround information in the same database.

The popularity of WAIS has been a great boost to the world of online information. Its integration with universal naming and hypertext is to be greatly encouraged.

Menu Systems and the Web

The Alex (Cate, 1992), Gopher (Alberti et al, 1991), and Prospero (Neuman, 1992) systems each use the directory and file (or menu and document) model to implement a global information system. These map into the web very naturally, as each directory (menu) is represented by a list of text elements linked to other directories or files (documents). These systems are very comfortable for readers who are used to hierarchical file systems, for whom directories are an established concept. Even when the structure is in fact cross-linked, readers feel at home as they regard it as a tree structure. Furthermore, for the information provider such systems are easy to build by cross-linking existing file systems.

> **Enthusiastic users of the browsing software particularly appreciated the consistent user interface for all types of data.**

An example of mapping a menu system onto the web is made by the W^3 client software which incorporates the simple Gopher protocol and therefore allows links into the Gopher system. The easy start-up of these systems has made them fairly popular. It is true that a menu is necessarily a more restricting medium of communication than general hypertext: a page of hypertext can convey more information to the reader about the choices to be followed, because

it uses more flexible formatting. Hypertext allows menus of links to lead to nodes with progressively greater textual content. However, the restricted world of plain text and menus, with its ease of publication, is adequate for many information providers.

Similarly, W³ clients also have built-in ability to browse the world of anonymous FTP archives, and a gateway provides access to Digital™'s VMS™/Help information.

X.500 and the Web

The x.500 standard for name servers provides a useful tool for long-term naming of documents. Initially intended for coordinates of people and organizations, to be used for documents it needs extensions similar to (though simpler than) those proposed, for example, by Yeong (1991). The chief attribute of a document for W³ purposes is the W³ physical address. Once access to x.500 name servers is widely available, "User Friendly Names" will form an appropriate W³ document name format for logical addresses.

Experience with the W³ Pilot Project

The first client software written to the W³ requirements ran on the NeXT machine using the NeXTStep™ graphic user interface tools. This hypertext browser/editor demonstrated the ease of use of a window-based hypertext interface to global information. It also allowed an overview hypertext database to be built and to point to data on the web by subject or organization. The second client written was a line-mode browser for character-mode terminals Being portable to almost any machine, it assures universal readability of all published documents. Hypertext documentation was put online, and gateways were set up into various existing information systems.

Enthusiastic users of the browsing software particularly appreciated the consistent user interface for all types of data. Reading news articles as hypertext is a good example: the same user interface is provided, and references between articles, and between articles and the news groups in which they are published, are all consistently represented as links.

It became evident that both hypertext links and text search were important parts of the model. A typical information hunt will start from a default hypertext page by following links to an index. A search of that index may return the required data, or some more links may be followed. Sometimes a further index may be found, and that searched, and so on. When the user of a hypertext editor has found what he wants (no matter how remote), he can make a new link to it from his home page so that he can find it again later almost instantly. This is generally preferable to making a copy that may soon be out of date.

The Future

The success of the pilot project prompted further development of W³-compliant software and information. Current client projects within various organizations include three X11-based browsers and a Macintosh browser. Various server gateways to other information systems have been produced, and the total amount of information available on the web is becoming very significant, especially since it includes all anonymous FTP archives, WAIS servers, and Gopher servers as well as specific W³ servers. We notice that a W³ server could provide the functions of each of these servers, and so we look forward to a single protocol that can be used by the whole community.

The Archie project (Emtage & Deutsch, 1992) provides an index into the Internet archives and is an excellent example of a service that we hope to make available in the web. We can imagine such indexing being extended to cover other forms of data. W³ provides a basic infrastructure for information access. All kinds of indexing, searching, filtering and analysis tools could usefully be built using the generic W³ access mechanism, and so be applied to all the various domains of data. Their results could then be made available on the web. Many possible research projects in hypertext are made possible by the existence of a very large linked information base.

Meanwhile, the W³ team at CERN and collaborators worldwide invite any information suppliers to join the web, contributing information or software. Detailed information about W³ protocols and data formats, and so forth, is available from our W³ server. The crudest way to access this is by Telnet to info.cern.ch. A better way is to run browser software (available by anonymous FTP from the same host) on your local machine. If you use a window-oriented browser, then you will be able to read articles like this on your screen. When you do, pick up your pen, mouse, or favorite pointing device and press it on a reference in this document.... The dream is coming true.

References

Alberti, Bob, Anklesaria, Farhad, Lindner, Paul, McCahill, Mark, & Torrey, Dan. (1991). *Notes on the Internet Gopher protocol.* Unpublished manuscript, Minneapolis: University of Minnesota, Microcomputer and Workstation Networks Research Center.

Beyond Hypertext: The DECWindows hyperenvironment vision. (1990). Maynard, MA: Digital Equipment Corporation.

Bush, Vannevar. (1945, July). As we may think. *The Atlantic Monthly,* 176 (1), 101–108.

Cate, Vincent. Unpublished private communication, January 27, 1992. Carnegie-Mellon University.

Davis, F.; Kahle, Brewster; Morris, Harry; Salem, Jim; Shen, Tracy; Wang, Rod; Sui, John; & Grinbaum, Mark. (1990). *WAIS interface Protocol: Protocol Functional specification.* Unpublished report. Menlo Park, CA: Thinking Machines Corporation.

Emtage, A., & Deutsch, D. (1992). Archie—An electronic directory service for the Internet. In *Proceedings of the Winter 1992 USENIX Conference* (pp. 92–110). Berkeley, CA: USENIX Association [2560 Ninth St., Suite 215, Berkeley, CA 94710].

Goldfarb, Charles F. (1991). *Information Technology-Hypermedia/Time-based structuring language (HyTime) (ISO/IEC CD 10744).* Unpublished manuscript [available from authors].

Halasz, F., & Schwartz, M. (1990). The Dexter Hypertext reference model. In Moline J.; Benigni, J.; and Baronas, J., eds. *Proceedings of the Hypertext Standardization Workshop January 16–18,* pp. 95–133. Gaithersburg, MD: National Institute of Standards and technology.

Kahn, Pau, & Meyrowitz. (1988). *Guide, HyperCard, and Intermedia: A comparison of Hypertext/Hypermedia systems.* (IRIS Technical Report 88-7). Providence, RI: Brown University.

Kantor, B. & Lapsley, P. (1986, February). A *proposed standard for the stream-based transmission of news* (Internet RFC-977).

Nelson, Theodor H. (1990). *Literary Machines version 90.1.* Sausolito, CA: Mindfull Press [3020 Brudgeway, Suite 295, Sausolito, CA 94965].

Neuman, Clifford B. (1992). Prospero: A tool for organizing Internet resources. *Electronic Networking: Research, applications, and Policy.* 2 (1), 31–38.

Postel, J., & Reynolds, J. (1985). *File transfer protocol (FTP)* Internet RFC-959).

Yeong, W. (1991) *Towards networked information retrieval* (Technical report 91-06-25-01). Reston, VA: performance Systems International, Inc.

Yeong, W. (1991, November). *Representing public archives in the directory.* Internet Draft. Reston, VA: Performance Systems International, Inc.

Wide Area Information Servers: An Executive Information System for Unstructured Files

Brewster Kahle, Harry Morris, Franklin Davis, Kevin Tiene, Clare Hart, and Robin Palmer

In this paper we present a corporate information system for untrained users to search gigabytes of unformatted data using quasi-natural language and relevance feedback queries. The data can reside on distributed servers anywhere on a wide area network, giving the users access to personal, corporate, and published information from a single interface. Effective queries can be turned into profiles, allowing the system to automatically alert the user when new data are available. The system was tested by twenty executive users located in six cities. Our primary goal in building the system was to determine if the technology and infrastructure existed to make end-user searching of unstructured information profitable. We found that effective search and user interface technologies for end-users are available, but network technologies are still a limiting cost factor. As a result of the experiment, we are continuing the development of the system. This article will describe the overall system architecture, the implemented subset, and the lessons learned.

Systems that allow corporate executives to access personal, corporate, and published information such as memos, reports, manuals, and news are new in the field of information management. The first integrated systems are just now coming on the market. They exploit networking, online mass storage, and end-user search systems; each of these has existed for some time, but their combination and integration has not been available for the corporate environment.

Commercial systems exist in each of the personal, corporate, and published data areas, with different levels of user friendliness. ON Location™, for instance, allows easy content-based retrieval of personal files on a Macintosh, whereas Lotus Magellean™ performs a similar function on a PC. Verity's Topic™ system allows for searching of LAN-based (usually corporate) archives but primarily for a trained user community. Dialog, Dow Jones, and Mead Data are major online providers of published information, but again the majority of their users are professionals in the field of information retrieval, such as corporate librarians.

Academic systems have also been developed for some of these applications. The Information Lens project (Malone, Grant & Turlack, 1986) uses structured electronic mail to help in automatic

Brewster Kahle *<brewster@think.com> is Project Leader for Wide Area Information Servers at Thinking Machines Corporation, 245 First Street, Cambridge, MA 02142. He has been with the company since 1983 when the company was founded. He was architect of the CPU of the Connection Machine Model 2 and led the design of the custom chips for the machine.*

Harry Morris *<morris@think.com> is a scientist on the Wide Area Information Servers (WAIS) project team, also at Thinking Machines. He developed the WAIStation (the Macintosh interface for WAIS), the Z39.50 library, and the disk-based Connection Machine search engine.*

Franklin Davis *<fad@think.com> is a project manager in the Advanced Information Systems Group at Thinking Machines, He developed software to handle huge databases on the massively parallel Connection Machine Document Retrieval System.*

Kevin Tiene *<tiene@apple.com>is a computer scientist in the Advanced Technology Group at Apple Computer, Inc. 20525 Mariani Blvd., Cupertino, CA 95014. He has been doing research and prototyping involving information retrieval, autonomous agents, and human interface. He received his computer science degree from George Mason University.*

Clare Hart *is Special Projects Manager for Dow Jones Information Services' Advanced Systems Group, 5918 Red Coat Lane, West Bloomfield, MI 48322. She was responsible for relations between Dow Jones and software developers known as Alliance Developers. Ms. Hart received a Bachelor of Science degree in Finance and Computer Systems Management from Drexel University.*

Robin Palmer *<palmer.r@applelink.apple.com> is a Senior Manager at Peat Marwick, 50 W. San Fernando Street, Suite 1200, San Jose, CA 95113 responsible for managing the systems integration microcomputer consulting group in San Jose. He has extensive audit and systems consulting experience for small, medium and large high technology and manufacturing companies. He has a Bachelor of Accounting, from the University of the Witwatersrand, Johannesburg, South Africa.*

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 59–68. © 1992 Meckler

organization and retrieval of business information. Project Mercury (Ginther-Webster, 1990) is a remote library searching system that uses a client-server model. The Smart system (Salton, 1971) is an information retrieval system that embodies many different searching strategies. The SuperBook project (Egan et al., 1989) is working on user interfaces for information systems, concentrating on the scientific user. Each of these systems is breaking new ground, but there is still no complete solution for the business executive wishing to search diverse information sources.

The Wide Area Information Servers (WAIS, pronounced "ways") system was constructed to test the acceptability of an integrated search system directly targeted at executives (Kahle, 1989). The companies participating in the project offered expertise in different parts of the problem: Dow Jones, with its business information sources; Thinking Machines, with its high-end information retrieval engines; Apple, with its user interface background; and KPMG Peat Marwick, with its information-hungry user base. Through this project, we wanted to determine if the wide area information retrieval market could incorporate users other than those trained searchers who are familiar with a variety of query languages and databases.

In the WAIS project we used a general architecture and built a small implementation to test the feasibility of an integrated information retrieval system for corporate end users. This article is a report on the overall architecture, the various implementations, and the lessons learned from this work.

The WAIS Architecture

The WAIS system took advantage of available technology to make a system that could then be tested on corporate executives to determine user acceptability. The system was composed of clients, servers, and the protocol that connects them. The information servers were Connection Machine systems, running a parallel signature-based search algorithm (Stanfill & Khale, 1986). The cross-country network connected several LANs with leased lines running AppleTalk and TCP, and carrying a variation on the Z39.50 application protocol. The clients ran on Macintoshes. This section describes the overall architecture, and the next section describes exactly what was implemented and used during the experiment.

The WAIS architecture was intended to have the following characteristics:

- Accessibility to novice users—little or no training should be required in order to perform effective searches.
- Remote accessibility—the servers must be accessible over a variety of networks.
- Uniform interface—a variety of databases, whether personal, corporate, or published, must be accessible from the same user interface.
- Automatic alerting—it must be easy to create profiles for background searching.
- Scalability—the system must scale in number of servers, size of servers, and intelligence of servers.
- Security—individuals and groups should be able to maintain control over who accesses their data.
- Flexible pricing model—a variety of information pricing structures, from per-minute charges to subscriptions, must be supported.
- Multimedia—the system must support the retrieval of any file format.

Many of these goals were achieved, while others, such as pricing model experimentation, were left unresolved.

In a client-server system, the client program is the user interface, the server does the searching and retrieval of documents based on indices, and the protocol (an agreed upon set of procedures) is used to transmit the queries and responses. The client and server are isolated from each other through the protocol so that they can be physically distant and interchangeable. Any client that is capable of translating a user's request into the standard protocol can be used in the system. Similarly, any server capable of answering a request encoded in the protocol can be used. In order to promote the development of both clients and servers, the protocol specification is in the public domain, as is its initial implementation.

On the client side, searches are formulated as quasi-natural language questions. The client application then formats the query for the WAIS protocol and transmits it over a network to a server. The server receives the transmission, translates the received packet into its own query language, and searches for documents satisfying the query. The ranked list of relevant documents are then encoded

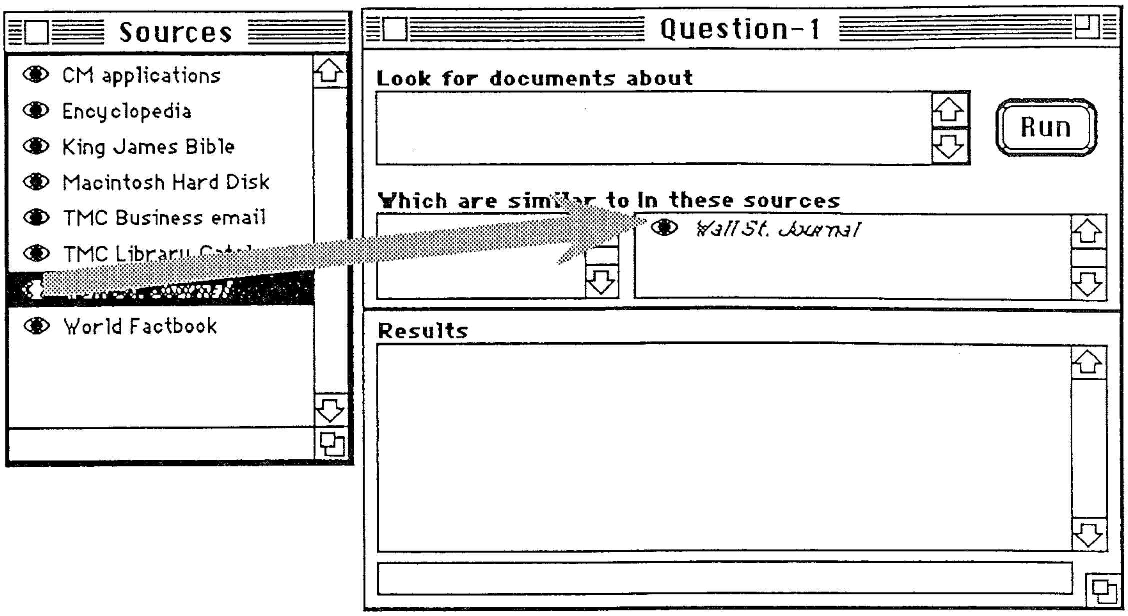


Figure 1. Sources are dragged with the mouse into the Question Window. A question can contain multiple sources. When the question is run, the program asks for information from each included source.

in the protocol and transmitted back to the client. At this point, the servers do not "understand" the quasi-natural language question posed by the user in any sense that a human would, but they use the words and phrases in the question to find documents that use those terms. The client decodes the response and displays the results. Documents of interest to the user can then be retrieved from the server.

Searching

We modeled the searching strategy on the interactive process people use when talking with a reference librarian. The library scenario is one in which the patron approaches a librarian or researcher with a description of needed information. The librarian might ask a few background questions, and then draw from appropriate sources to provide an initial selection of articles, reports, and references. The patron sorts through this selection to find the most pertinent documents. With feedback from these trials, the researcher can refine the search and even continue to supply the patron with a flow of information as it becomes available. Monitoring which articles were retrieved can help the researcher provide appropriate information for future searchers.

The WAIS system uses a similar means of interaction: the user states a question in unrestricted natural language to a set of sources, and a set of document descriptions is retrieved (see Figure 1). The server assigns each document a score, based on how closely the words in the document matched the question (see Figure 2). The user can examine any of the documents, print them, or save them for future use (see Figure 3). If the initial response is incomplete or somehow insufficient, the user can refine the question by stating it differently.

Once a relevant document is found, the user may say "I want more like this one" by marking the retrieved documents as being "relevant" to the question at hand, and then re-running the search (see Figure 4). This method of query refinement is called relevance feedback (Salton & McGill, 1983). The server uses the marked documents to attempt to find others that are similar to them. In the present WAIS server, "similar" documents are those that share a large number of statistically significant words and phrases. This brute force method works surprisingly well with large collections of documents (Stanfill, 1991; Stanfill & Khale, 1986).

A Common Protocol for Information Retrieval

One of the most far-reaching aspects of this project was the development of an open protocol. The four companies involved jointly specified a standard protocol for information retrieval by extending an existing public standard, Z39.50-1988 (NISO, 1988). We chose this public standard rather than inventing one ourselves since it was close to what we needed

and it could help us keep the protocol from being regarded as proprietary.

The use of an open and versatile protocol can foster hardware independence and competition. This not only provides for a much wider base of users, but it also allows the system to evolve over time as hardware technology progresses. For example, the protocol provides for the transmission of audio and video as well as text, even though at present most personal computers are unable to handle such transmissions. However, computers are free to ignore pictures and sound returned in response to questions, and to display and retrieve only text, if that is all they are capable of processing. Higher end platforms are free to exploit their greater processing power and network bandwidth.

Z39.50 is a general attribute-based Boolean search protocol intended to run over the Open Systems Interface (OSI) stack. It was designed for search and retrieval of bibliographic Machine-readable Cataloging (MARC) records in libraries. As such, its structure allows easy access to traditional Boolean search systems such as STAIRS (Salton & McGill, 1983).

The WAIS protocol is an extension of the existing Z39.50-1988 standard, but we are working with the standards committee to merge the extensions back into the newer versions (Davis, et al., 1990). The extensions allow support for multimedia data, large documents, a directory of servers, different communication systems, and distributed retrieval. To support multi-media, a document must be available in a number of formats. This was accomplished by listing the set of available types in the search response from which the client can choose one to retrieve. Another problem with the protocol involved retrieving large records. Large documents, text or nontext, would be slow to display if the whole document had to be retrieved at one time, as is required in the original standard. Large documents are supported in the WAIS protocol by allowing the client to retrieve sections of a document based on the number of bytes or lines requested.

We also standardized a format for describing servers (Kahle & Morris, 1991a) and how to contact them, which is necessary to implement a directory of servers. To support communication systems other than the full OSI protocol stack, a header was needed to show how long the packet was and how it was encoded. With this packet header we implemented the WAIS protocol over modems, TCP/IP, and X.25 systems. To support distributed retrieval we needed a document identifier system that could be used in a distributed environment (Kahle & Morris, 1991b).

The protocol used in the WAIS system has proven useful in the distributed full-text environments in which we tested it.

User Interfaces: Asking Questions

Users interact with the WAIS system through the Question interface. Each question form has an area for the user's quasi-natural language question, the list of sources that will be accessed to try to answer the query, the list of relevant documents, and a list of answer documents.

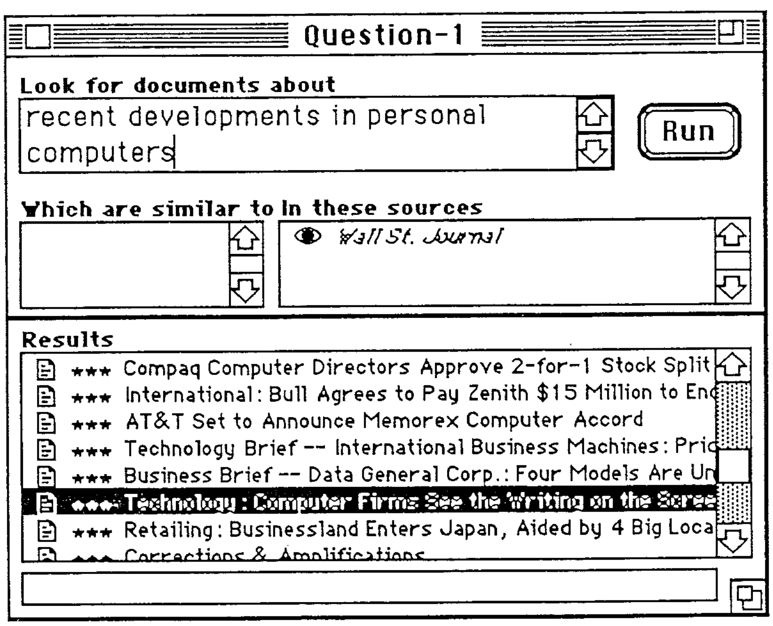


Figure 2. When a query is run, headlines of documents matching the query are displayed.

The illustrations here are taken from the initial WAIStation program produced at Thinking Machines for the Apple Macintosh. We have also built clients for X windows and gnu-emacs. Another Macintosh interface was developed that emphasizes the alerting feature (Erickson & Salomon, 1991).

With most current retrieval systems, complications develop when one begins dealing with more than one source of information. For example, one contacts the first source, asks it for information on some topic,

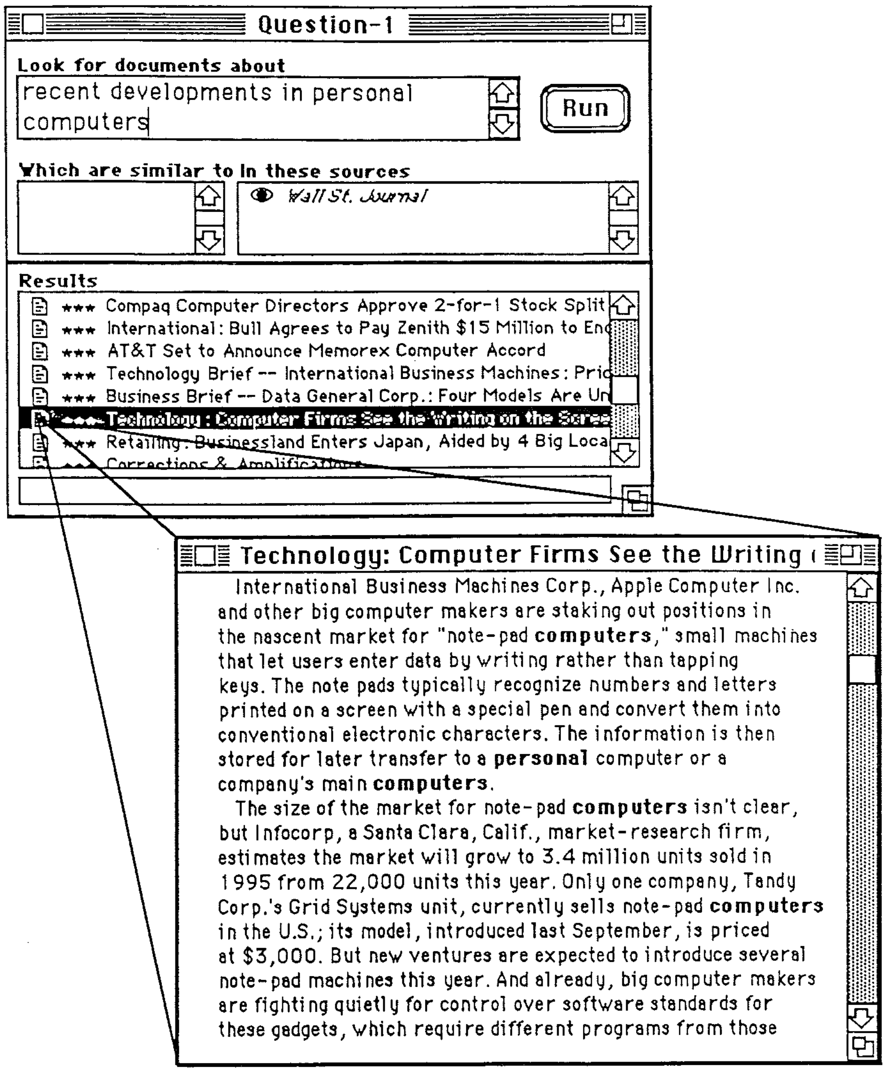


Figure 3. With the mouse, the user double clicks on any resulting document to retrieve it. The document can contain graphics.

be manipulated by the user. The user has transparent access to a multitude of local and remote databases.

From the user's point of view, a server is a source of information. It can be located anywhere: on the local machine, on a network, or on the other side of a modem. The user's workstation keeps track of a variety of information about each server. The public information about a server includes how to contact it, a description of the contents, and the cost. In addition, individual users maintain their own private information about the servers they use.

Users may need to budget the money they are willing to spend on information from particular servers, know how often and when each server is contacted, and assess the relative usefulness of each server. In the current interface, the budget entries were put in as placeholders, since all servers are currently free. When a source is contacted, all questions that refer to the source are updated with the new results.

contacts the next source, asks it the same questions (most likely using a different query language, a different style of interface, and a different system of billing), contacts the next source, and so on. One of the primary goals behind the development of the WAIS system was to replace all this with a single interface.

With WAIS, the user selects a set of sources to query for information and then formulates a question. When the user presses the RUN button (see Figure 2), the system automatically asks all the desired servers for the required information, with no further interaction necessary by the user. Thus, the documents returned are sorted and consolidated in a single place, to

A "confidence factor" allowed users to multiply the score returned from different servers so that the list presented to the user would be more appropriate. This was put in the interface to anticipate a number of different server technologies with different scoring algorithms. The "confidence factor" allows the user to adjust the scores. In addition, a user might have a preference for the information from one server over another, so a subjective balance would be helpful. This feature was rarely, if ever, used because the number of servers was small, they all used the same server technology, and most users only asked one source at a time.

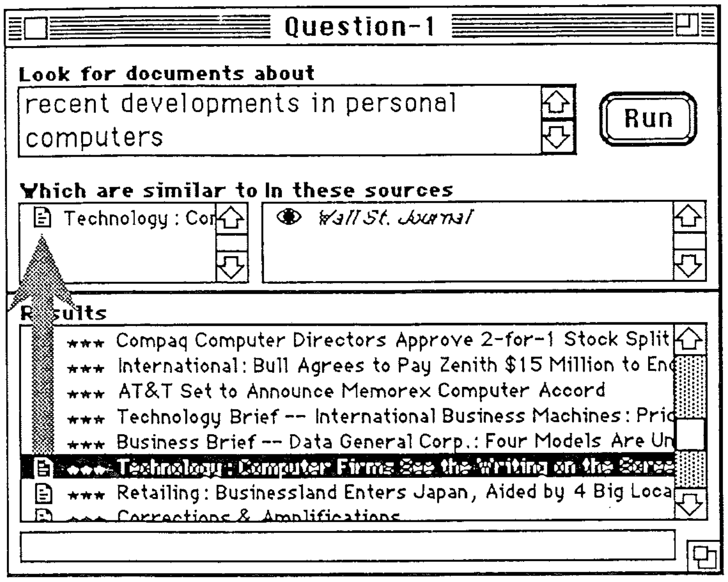


Figure 4. To refine the search, one or more of the result documents can be moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents that are "similar" to the ones selected.

Servers

The servers in the WAIS system hold databases that can be queried by a client. References of documents that best match the words and phrases in the query are returned to the client. A client can then request all or part of a document from the server. Since the client explicitly contacts the server, any number of billing methods can be employed such as 900 numbers, credit cards, and subscriptions.

The Connection Machine server system (CMDRS), used in the WAIS system, stores the documents in a compressed form, called signatures, which can be searched quickly using the parallel processors of the Connection Machine (Stanfill & Khale, 1986). The signatures are stored in the RAM of the machine thereby assigning a few documents to each processor of the machine. Each word in the query is then broadcast to all the processors, and a score is kept for each document to reflect the number of words and phrases that matched. Weighting is done based on crude proximity and occurrence frequency. The resulting search results have been found to be useful to end-users.

As the dissemination of information becomes easier, questions of ownership, copyright, and theft of data must be addressed. These issues confront the entire information processing field, and are particularly acute here. The WAIS system is designed to keep control of the data in the hands of the servers. A server can choose to whom and when the data should be given. Documents are distributed with an explicit copyright disposition in their internal format. This is not to say that theft cannot occur, but if a client starts to resell another's data, standard copyright laws can be invoked. By keeping the control of the distribution of works with the creators, many of the problems of copyright do not arise.

Rerunning Questions— A Personal Newspaper

In addition to providing interactive access to information, the WAIS system can also be used as a rudimentary personal newspaper to alert its user when new documents are available on a subject that might be of interest (see Figure 5). In the library literature, this is referred to as selective dissemination of information (SDI), and many manual, semi-automated, and automated systems have been implemented. Our initial implementation involves saving interactive questions and automatically rerunning them at periodic intervals, checking if new documents are available. This technique has the advantage of hiding communication costs, using systems off hours, and finding potentially interesting information in a timely manner.

Multimedia Database

The documents retrieved through WAIS may be any kind of file, such as text, still graphics, motion pictures, or hypertext documents. The searching of the system is based on an initial quasi-natural language question and further relevance indications, but the server is free to use that information in any way to find appropriate documents. The protocol simply defines a document as a block of data and a type. The client uses the type to determine how to display the document. A list of available types is part of the search response of each document. This allows clients to choose among a selection of types and suppress documents whose types they cannot display. Alternatively, they can simply store the documents in their local disk for latter processing.

Our initial X windows clients are able to use other programs to display graphic data such as Tagged Image Field Format (TIFF) and Graphics Interchange Format (GIF). The Macintosh client can display PICT images and text, but can theoretically download any type of file.

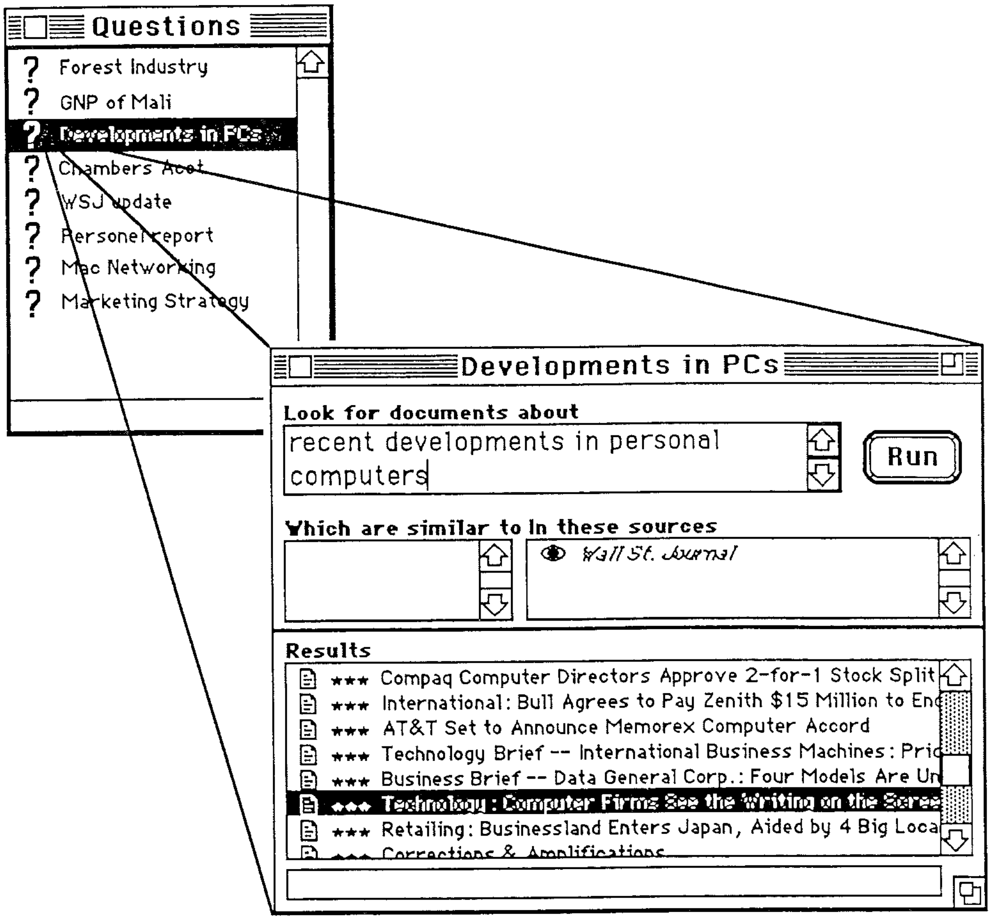


Figure 5. Opening a saved question which was automatically updated in the background and contains new data.

Nontextual data are indexed in one of two ways. If the data include an embedded description (e.g., TIFF), the description is used for indexing. Otherwise an external description is indexed. When a search identifies the description file as a suitable response, the multimedia data are returned instead of the description file.

The Directory of Servers

To find sources of information in a distributed environment, we used a "directory of servers" which is a database of documents describing other servers. In response to a query, the database of servers is searched, returning a list of documents (i.e., server descriptions) that match the query. Instead of text documents, however, it takes advantage of the mixed type capabilities of WAIS to return a structured document with many specific fields for cost and contact information (see Figure 6). This capability will become more important as the number of servers increases.

For example, suppose you needed information concerning the current gross national product of Mali but had no idea on which server to find it. You could first ask the directory of servers for "information about the current economic condition of Mali." The directory will take the words in the query and find descriptions of the servers that contain those words. It might then return several documents. The *World Factbook*, for instance, might appear because of a match on "economic condition." This source description could then be used as the source field of another question. This time the system would contact the *World Factbook*, ask for the information, and possibly return a document with a description of Mali (*World Factbook*, 1974).

In addition, the directory of servers provides a means for information providers to advertise the availability of their data. When a new source becomes available, the developers can submit a textual description, along with the necessary information for contacting the server. This information is added to the directory and becomes available to the public by the searching interface.

The Prototype WAIS System

In the fall of 1990 we installed an experimental WAIS system at Peat Marwick. The prototype was used by 20 users in six cities. Peat Marwick utilized corporate data in Montvale, New Jersey, and Dow Jones information in Princeton, New Jersey. The

system was run successfully for six months with good user reactions.

KPMG Peat Marwick is an example of an information-intensive company. Their role as consultants requires that they maintain an awareness of new products, market fluctuations, changing laws, internal regulations, and competition. In addition, as a large organization, it possesses considerable internal information, such as company contacts, bids, reports, and resumes. Furthermore, distributing such material in forty countries, with 200 offices in the United States alone, makes the company a prime candidate for wide area information technology.

The primary users were located in San Jose and connected by 56kbaud and 9.6kbaud circuits to the servers in New Jersey. The 20 managers and partners in the Peat Marwick's accounting division used an 8192 processor Connection Machine system for serving reports; proposals; resumes; contracts, accounting manuals; the Peat Marwick Audit Manual, Management Guide, and Professional Development Courses; documents from the Financial Accounting Standards Board, the Government Accounting Standards Board, and the American Institute of Certified Public Accountants; and a tax library. The data were separated into twelve different databases which could be searched separately or in any combination. There was also a virtual database consisting of all these sources.

The connection to Dow Jones provided access to 1 gigabyte of data, running on a 32K processor Connection Machine. The data consisted of a year of the *Wall Street Journal, Barron's*, and 400 magazines. Each of the approximately 250,000 articles was a separate document. The ability to search personal data was not available at the time of the experiment.

Lessons Learned

The search technology performed well in finding useful data for end-users who were given little instruction about system use. The speed of the searches (usually between two and ten seconds) depended on the communication speed, since the search itself took much less than a second. When the response time was greater than 10 seconds, the users voiced complaints, but in general they were very pleased with the search results. The ability to execute searches without prior training and without in-depth knowledge of the database was essential to the users. Relevance feedback was used frequently and effectively by users who were aware of its existence. Not all users realized it was available, however. This is an opportunity for improving user interfaces. For example, relevance feedback could be performed automatically on any document the user chooses to view. This would result in a kind of automatic, dynamically linked hypertext system, where every document is "linked" to all similar documents.

The Macintosh user interface (WAIStation) also performed well in terms of ease of use and adaptability. With a single demonstration, most users were able to execute searches and save their results. Left with only the manual, new users took 15 to 30 minutes to feel comfortable with the system. The ability to search local and remote databases transparently was greatly appreciated, as reported in user feedback forms. The biggest problem we had with the interface was in implementing the TCP and modem connections from the Macintosh. The automatic updating feature of WAIStation was rarely used and needs more work to make it more obvious and to allow it to give better feedback when documents are found.

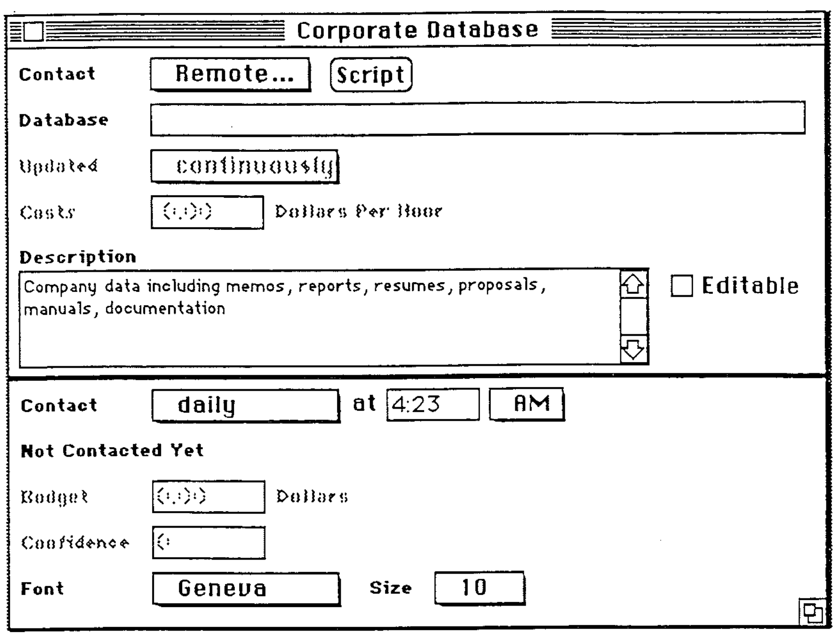


Figure 6. The Source description contains all the necessary information for contacting an information server.

Wide area communications proved to be a difficult part of the project due to our resistance, based on future cost projections, to use leased lines. The original plan called for linking San Jose and Montvale with Shiva Telebridges™ running at 9600 baud on a normal phone line. This approach did not prove reliable, nor did it give us reasonable performance. We ended up replacing this link with a dedicated 56kbaud line attached to a SyncRouter (Engage Communications™). The dedicated line was highly reliable, and 56kbaud was fast enough to support many active users of the system, while maintaining an interactive feel in both search and retrieval.

Organizing and formatting the data for display on the client workstation proved to require more effort than we expected. The current Macintosh client is capable of displaying only ASCII text and PICT format picture files. This meant that the corporate data, which consisted primarily of word processor files, had to be converted to ASCII. Since the conversion was not perfect, some documents required a small amount of manual reformatting. This is obviously unacceptable in a production system. A more attractive solution might be to build a client that can display the most common document formats and that can call on other applications to display formats it cannot understand. This approach will become easier to implement as document filters (e.g., Claris™ XTND) and interprocess communication become more common. This approach will also make it possible to index and store the original document rather than an ASCII shadow.

As the searchable Peat Marwick corporate collection grew, the users wanted to search just parts of the database. The natural divisions for the users were the original sources of the text, such as training manuals or government legal texts.

In summary, we found that the users were pleased with the system, and some used it many times each day. It appears that there is a market for end-user search systems and that the technology is ready. The weak link seems to be communication infrastructure.

Conclusion

In developing the WAIS system, the participating companies have demonstrated that current hardware technology can be used effectively to provide sophisticated information retrieval services to novice end users. How this might affect information providers is not yet understood. The users at Peat Marwick found the technology useful for day-to-day tasks such as researching potential new accounts and finding resources within their own organization. Since these tasks are not restricted to the accounting and management consulting industries, we are optimistic that this type of technology can be fruitful and productive in many corporate settings.

The future of this system, and others like it, depends on finding appropriate niches in the electronic publishing domain. Potential uses include making current online services more easily accessible to end-users and allowing large corporations to access their own internal data more effectively. It is also possible that near-term development will focus on a single professional field such as patent law or medical research.

Acknowledgments

The design and development of the WAIS Project has been a collective effort, with contributions and ideas coming from many people. Among them are the following:

- Apple Computer: Charlie Bedard, David Casseras, Steve Cisler, Ruth Ridder, Eric Roth, John Thompson-Rohrlich, Kevin Tiene, Gitta Soloman, Oliver Steele, Janet Vratny-Watts.
- Dow Jones News/Retrieval: Rod Wang, Roland Laird.
- KPMG Peat Marwick: Chris Arbogast, Mark Malone, Tom McDonough.
- Thinking Machines: Dan Aronson, Patrick Bray, Jonathan Goldman, Danny Hillis, Rob Jones, Barbara Lincoln, Gordon Linoff, Chris Madsen, Gary Rancourt, Sandy Raymond, Steve Schwartz, Tracy Shen, Craig Stanfill, Robert Thau, Ephraim Vishniac, David Waltz, Uri Wilensky.

References

Davis, F., Kahle, B, Morris, H., Salem, T, Shen, T, Wang, R., Sui, J., & Grinbaum, M. (1990, April). *WAIS interface protocol prototype functional specification*. Unpublished paper. Menlo Park, CA: Thinking Machines. (Available via anonymous ftp: /pub/wais/doc/wais-concepts.txt@quake.think.com OR wais server wais-doc.src.)

Egan D., Remde, J., Gomez, L., Landauer, T., Eberhardt, J., & Lochbaum, C. (1989). Formative

design-evaluation of SuperBook. *AMC Transactions on Office Information Systems, 7*, 30–57.

Erickson, T., & Salomon, G. (1991). Designing a desktop information system: Observations and issues. In S. P. Robertson, G. M. Olson, & J. S. Olson (Eds.), *Human Factors in Computing Systems: Reaching Through Technology, CHI'91 Conference Proceedings* (pp. 49–54). New York: ACM.

Ginther-Webster, K. (1990). Project mercury. *AI Review of Products, Services and Research, 3* (July-August), 25–26.

Kahle, B. (1989, November). *Wide area information servers concepts* Menlo Park, CA: Thinking Machines. Thinking Machines technical report TMC-202. (Available via anonymous ftp: /pub/wais/doc/wais-concepts.txt@quake.think.com. OR wais server wais-docs.src.)

Kahle, B., & Morris, H. (1991a, February). *Source description structures.* Menlo Park, CA: Thinking Machines. (Available via anonymous ftp: /pub/wais/doc/source.txt@quake.think.com.)

Kahle, B., & Morris, H. (1991b, May). *Document identifiers or international standard book numbers for the electronic age.* Menlo Park, CA: Thinking Machines. (Available via anonymous ftp: /pub/wais/doc/doc-ids.txt@quake.think.com.)

Malone, T., Grant, K., & Turback, F. (1986). The information lens: An intelligent system for information sharing in organizations. *Human Factors In Computing Systems, Special Issue of the SIGCHI Bulletin C.H.I.'86 Conference Proceedings* (pp. 1–8). New York: ACM.

NISO. National Information Standards Organization. (1988). *Z39.50-1988: Information retrieval service definition and protocol specification for library applications.* National Information Standards Organization (Z39), P.O. Box 1056, Bethesda, MD 20817. Telephone (301)975-2814. (Available from Document Center, Belmont, CA. Telephone (415)591-7600.)

Salton, G. (1971). *The SMART retrieval system—experiments in automantic document processing.* Englewood Cliffs, NJ: Prentice-Hall.

Salton, G., & McGill, M. (1983). I*ntroduction to modern information retrieval.* New York: McGraw-Hill.

Stanfill, C. (1991). *Massively parallel information retrieval for wide area information servers.* Unpublished paper presented at the International Conference on Systems, Man, and Cybernetics, Charlottesville, Virginia.

Stanfill, C., & Kahle, B. (1986). Parallel free-text search on the connection machine system. *Communications of the ACM, 29*, 1229-1239.

World Factbook. (1974). Acton, MA: Publishing Sciences Group.

Books of Interest

Dial In: An Annual Guide to Library Online Public Access Catalogs

Michael Schuyler

This directory lists the dial-in numbers to online public access catalogs (OPACS) from hundreds of libraries internationally. Entries include library name, address, data on special collection strengths, network membership, Internet addresses, loan policies, requirements, and restrictions on access.

ISSN 1047-3424
$55.00 paper ISBN 0-88736-808-5 1992
250pp. Published annually in November
5% Standing Order Discount Available

Search Sheets for OPACs on the Internet: A Selective Guide to U.S. OPACs Utilizing VT100 Emulation

Supplement to Computers in Libraries, Number 32
Marcia Klinger Henry, Linda Keenan, and Michael Reagan

Search sheets summarize the information a user needs to know to access library online public access catalogs (OPACs). This volume includes samples of search sheets from a number of the major OPAC software systems now in use. Also included are a survey of log-on and search techniques and an appendix of sample help screens to the systems discussed.

$39.50 ISBN 0-88736-767-4
200pp. 1991

Libraries, Networks and OSI: A Review, with a Report on North American Developments, 1992 edition

Supplement to Computers in Libraries, Number 49
Lorcan Dempsey

This report (published in association with the United Kingdom Office of Library Networking) details a study trip to North America to analyze the use of computer networks by libraries. Coverage includes the computer networking context in the UK, Europe, and North America.

This study describes for the first time the impact and implications of computer networks and networked information resources on librarians, campus information technology planners, and those concerned with national networking policy.

$49.50 paper ISBN 0-88736-818-2
240pp. January 1992

Local Area Networks in Libraries

Supplement to Computers in Libraries, Number 27
Kenneth Marks and Steven Nielsen

Introduces the subject of microcomputer- and minicomputer-based local area networks within the library environment. All aspects of the topic are covered, from the most elementary to a discussion of complex integrated networking systems.

$42.50 ISBN 0-88736-705-4
200pp. 1991

Using Computer Networks on Campus

Papers from the 1st Annual Conference (1990)
Edited by Les Lloyd

These proceedings from the first conference on Campus-Wide Information Systems (CWIS), held at Lafayette College (Easton, PA) in June 1990, focus on key issues in the design, implementation, and management of electronic information networks in an academic setting.

$30.00 ISBN 0-88736-813-1
175pp. 1991

Using Computer Networks on Campus 2

Papers from the 2nd Annual Conference (1991)
Edited by Les Lloyd

The proceedings from the second annual conference on this topic include papers on intra-campus telecommunications, the role of the library in the campus network, network resources for academics and administrators, and other issues of concern to colleges and universities implementing electronic information networks.

$30.00 ISBN 0-88736-814-X
175pp. 1991

New in 1992

Advances in Online Public Access Catalogs, Volume 1

Edited by Marsha Ra

The annual publication offers a wide-ranging and stimulating discussion of electronic library catalogs mounted for public access. Issues of access to information, electronic bibliographic control, search software and protocols, are among the topics discussed.

$55.00 ISBN 0-88736-775-5
200pp. March 1992

Building Blocks for the National Network: Initiatives and Individuals

Supplement to Computers in Libraries, Number 34
Nancy Melin Nelson

Profiles, interviews, and analytical studies of the major individuals, issues, and organizations involved in the evolution of electronic information networking.

$35.00 ISBN 0-88736-769-0
160pp. July 1992

Campus-Wide Information Systems: Case Studies

Supplement to Computers in Libraries, Number 56
Edited by Les Lloyd

More and more colleges and universities are providing computing services to classrooms, faculty and administrative offices, off-campus buildings, and dormitories. In this volume, the editor, a director of academic computing services, gathers case studies from a number of different insitutions about their efforts to design and implement campus-wide information networks. Covered are such topics as hardware and software selection, access to networked information, mounting of local data files, telecommunications, evaluation of service, troubleshooting, and other issues of concern within the campus networking environment.

$30.00 ISBN 0-88736-834-4
250pp. May 1992

See other side for more Meckler titles ⇨ ⇨ ⇨

Directory of Computer Conferencing in Libraries

Supplement to Computers in Libraries, Number 36

Brian Williams

This volume is at once an introduction to computer-based conferencing (here defined as interactive electronic communication among remote terminals on a specific topic), a survey of the major electronic conferencing software, and a guide to the main conferencing systems that utilize that software. Discussion includes electronic bulletin boards, online educational systems, electronic mail systems, and public computer networks, such as the Internet, Tymnet, and Bitnet.

$65.00 ISBN 0-88736-771-2
500pp. January 1992

Directory of Directories on the Internet

Supplement to Computers in Libraries, Number 33

Ray Metz

The Internet, the high-capacity network of several hundred libraries and other institutions linked in a common electronic communications system, contains thousands of discrete forums of special interest information resources. This guide helps in navigating the Internet's maze of data and allows access to resources of interest to researchers and librarians.

$29.50 paper ISBN 0-88736-768-2
175pp. June 1992

Directory to Fulltext Online Resources 1992

Supplement to Computers in Libraries, Number 55

Jack Kessler

This book is an introduction to and a directory of fulltext resources currently available online. Among the areas and types of materials covered are commercial online services such as Dialog, BRS, Lexis, Nexis, etc, and CD-ROM fulltext databases. In addition, such specific fulltext projects as the Oxford Text Archive (a directory of fulltext documents produced in the U.K.), its U.S. counterpart, Project Gutenberg, and many other international efforts are discussed in detail. Many online public access catalogs (OPACs), electronic conferences, and electronic bulletin boards also contain fulltext material. This Directory will make these resources more widely known and accessible to researchers and librarians.

$30.00 paper ISBN 0-88736-833-6
140pp. February 1992

Electronic Information Networking

Supplement to Computers in Libraries, Number 45

Nancy Melin Nelson and Eric Flower

This volume collects presentations from the first Research and Education Networking Conference held in Oakland, CA, March 7-8, 1991. Contributions include: Research and Education Networks: Technical, Institutional, Human, and Political Perspectives; Community Access to National Networks; Navigating the Resources on the Networks; Canada and the North, International Connections; and Perspectives on Networking, Reaction to the Issues.

$35.00 ISBN 0-88736-815-8
165pp. April 1992

From A to Z39.50: A Networking Primer

Supplement to Computers in Libraries, Number 31

James J. Michael

An introduction to and discussion about the issues and standards involved in electronic telecommunications and the high-speed, high-capacity transfer of electronic data files.

$29.50 paper ISBN 0-88736-766-6
165pp. April 1992

Networked Information: Issues for Action

Supplement to Computers in Libraries, Number 46

Edited by Elaine M. Albright

This volume collects papers delivered at the ACRL New England Chapter Spring Conference (March 1991, Bowdoin College, Maine) on the topics of national information policy, the construction of a national electronic research network, local campus networking connections to the national system, and the role of libraries in a networking environment. Indexed.

$42.50 ISBN 0-88736-823-9
160pp. June 1992

Meckler

MANAGING INFORMATION TECHNOLOGY

Order Form

Meckler • 11 Ferry Lane West • Westport, CT 06880 • 203-226-6967
Meckler Ltd • 247-249 Vauxhall Bridge Rd. • London SW1V 1HQ • 071-931-9985

ELN92

Enter my order for the following books:
(10% discount for pre-payment, also, additional 10% off for subscribers)

Title ▼ Price ▼

____

____

____

Name ______________________ Organization ______________________

Address ____

City ______________ State/Province ______________ Zip/Postal Code ______________

❑ Bill Me ❑ Charge my ❑ VISA ❑ MasterCard ❑ AMEX

Account # ______________ Signature ______________ Phone # ______________

The Internet Gopher: An Information Sheet*

What is the Internet Gopher?

The Internet Gopher is an information distribution system. It combines features of electronic bulletin board services and databases, allowing you to either browse a hierarchy of information, or to search for the information you need using full-text indexes. Gopher can also store references to public telnet sessions, CSO phone book servers, finger-protocol information, Archie servers, WAIS servers, ftp sites, and sounds.

The Internet Gopher software was developed by the Computer and Information Services department of the University of Minnesota. The software is freely distributable.

What's Available?

There is a diverse collection of information stored on various Gopher servers: computer documentation, phone books, news, weather, library databases, books, recipes, etc. Since you can seamlessly navigate between servers with Gopher, you do not need to worry about exactly where a given piece of information resides.

We use Gopher at the University of Minnesota's Microcomputer Helpline to answer questions using our user support Q&A database (containing over 7000 Q&A items). While this is a good tool for our consultant's use, it is more important that users can directly access this database. This means fewer calls to our helpline, resulting in better, faster service.

The Gopher system can keep track of campus phone book servers. Currently you can search seventeen university phone books.

Quite a bit of news is in Gopher. There are two gopher-ized campus newspapers: both the Minnesota Daily and The Daily Texan are on line and searchable. National Weather Forecasts for the entire nation are also available. The University of Minnesota has a site license for the Clarinet UPI news service; we provide on-campus users with a full UPI news feed that's full-text indexed hourly.

The electronic books published by the Gutenberg Project are available in Gopher. These include classics such as Moby Dick and reference works such as the CIA World Fact Book.

There are also gateways between Gopher and Archie, ftp and WAIS servers so that a Gopher user can access items from many different sources without learning a new user interface for each system.

Gopher users can access information that is only accessible on terminal based information systems. Gopher can store links to these sites. You can easily start a telnet session to many libraries and information servers with the press of a key or click of the mouse.

---

*Editor's Note: This information sheet, provided by the Gopher development team, briefly describes the Gopher service and its use.

How does Gopher work?

Information is stored on multiple servers, connected together in a network. This allows for capacity to be added to the system in small, inexpensive increments. It also allows the Gopher system to cross institutional boundaries, since other servers can be "linked" into the system easily. Large indexes can be spread over multiple servers, resulting in significant speed ups.

You may use the Macintosh, PC, NeXT, VMS, VM/CMS, X-windows, or Unix terminal clients to access the Gopher system. The client connects with a "root" Gopher server which is an entry point into the Gopher hierarchy. There can be many different entry points. This allows a certain amount of freedom in organizing the information. Local or frequently accessed information can be put higher in the hierarchy for different organizations (i.e. the Library root server would have a library search at the top level, whereas the Music root server would have it lower)

At the initial connection, the root server sends back a listing of the objects in its top level directory. These objects can be:

Directories
Text Files
CSO Phone Books
Search Engines (Gopher, WAIS, Archie)
Telnet References
Sounds

Each object has associated with it a name to display to the user, a unique "selector string" to retrieve the object from the server on which it resides, a server hostname, and a port number. Given a list of objects, the Gopher client can present the list to the user, and the user can then make a selection. The user does not have to remember hos names, ports, or selector strings. The client takes care of this.

After the user makes a selection, the client contacts the given host at the given port and sends the selector string associated with the object. The client will respond differently, depending on what type of object was selected. The client may display a new directory, show a text file, or prompt the user to search a CSO phone book. This process continues until the user decides to quit.

Since Gopher uses a simple protocol, we and others were able to develop clients and servers on many platforms quickly and easily.

How do I access Gopher?

Client software for Macintoshes, PCs, NeXTs, X Windows, VMS, VM/CMS and UNIX terminals is available for anonymous ftp from

boombox.micro.umn.edu

in the directory

/pub/gopher

Or, if you just want a quick look at the UNIX terminal client, telnet to the machine

consultant.micro.umn.edu

and log in as:

gopher

We highly recommend running the client on your local personal computer or workstation. These local clients have a better response time and an easier user interface.

Contacting Gopher People

The University of Minnesota Gopher Development Team (Mark McCahill, Farhad Anklesaria, Paul Lindner, Bob Alberti, Daniel Torrey) can be reached by sending internet e-mail to:

gopher@boombox.micro.umn.edu

or by snail mail :

Gopher Preject
Computer and Informantion Services
Room 125 Shepedrd Labs
University of Minnesota
100 Union Street SE
Minneapolis, MN 55455

phone: (612) 625-1300
FAX: (612) 625-6817

You can subscribe to the Gopher-news mailing list by sending a request to:

gopher-news-request@boombox.micro.umn.edu

There is also a USENET newsgroup (alt.gopher) where Gopher is discussed.

Resource Reviews

Joe Ryan

Baum, Michael S. & Perritt, Henry H., Jr. (1991). *Electronic contracting, publishing and EDI law.* New York: John Wiley & Sons, Inc. 871 pp. Available: John Wiley & Sons, 1 Wiley Drive, Somerset, NJ 08873-1272. Phone: (908) 469-4400. ISBN: 0-471-53235-9

Wright, Benjamin. (1991). *The law of electronic commerce. EDI, fax and e-mail: Technology, proof, and liability.* Boston: Little, Brown & Company. 432 pp. Available: Little, Brown, & Company, 200 West Street, Waltham, MA 02254. Phone: (800) 343-9204 ISBN: 0-316-95632-5.

The technology that will permit parties to enter into contracts through the use of electronic networks by means of data communication protocols such as electronic data interchange (EDI) is with us today. These EDI protocols will make possible the elimination of the paper and the human signatures historically associated with commerce in goods and services. Standardized electronic analogues to the familiar written purchase orders, bills of lading, and invoices have been developed. This has enabled firms to communicate all the essential contract terms—quantity, price, delivery, and so on—necessary for the "meeting of the minds," which is central to the law of contract formation.

The use of confidential codes, encryption, or other security techniques can provide the parties to electronic business transactions with the requisite assurances that they are not dealing with an impostor and that the data being exchanged are authentic—in essence an "electronic signature." Even the proverbial "battle of the forms" by which buyers and sellers traditionally bombard each other with often conflicting "fine print" can be addressed through the use of "trading partner" or "interchange" agreements by which the parties settle in advance on the general conditions governing their transactions.

Ever since 1677, however, when the English Parliament passed an Act for the Prevention of Fraud and Perjuries, most contracts for the sale of goods were not enforceable unless evidenced by "some note or memorandum in writing of the said bargain...signed by the parties to be charged by such contract." This "statute of frauds" was intended not only to obviate perjury, but also to promote the public policy interest of certainty in the business transactions that undergird a nation's economy. This certainty in transaction was accomplished by adopting the most advanced data transmission and storage technology available at the time: paper and ink. Versions of this "statute of frauds" are the law in each of the states in the U.S. today as part of the Uniform Commercial Code.

Do such requirements for "writings" and "signatures" render contracts facilitated by EDI standards over electronic networks unenforceable? If not, is it possible to have electronic "documents" admitted as evidence in courts of law to prove the terms of a business transaction? If so, how can business relationships be structured so as equitably to allocate the risks associated with these new methods of communicating? What is the role of telecommunications networks themselves as service providers to the parties to these relationships? What are the implications of electronic networking for government policies regarding the creation and dissemination of information? Finally, what are the impacts on the law of intellectual property, including patents, copyrights, and trade secrets?

The answers to these and many other questions regarding the interface between the law and electronic networking technologies can be found in the two recent treatises cited above. These useful reference works are the first of their kind. By presenting the detailed analysis to be expected from traditional legal texts in the context of practical discussions of communications technologies and applications, they fill a serious need. Their basic message is simple and positive: there are no fundamental legal principles that should stand as a barrier to doing business electronically. Rather, certain basic legal principles, such as the relationship between principal and agent, are essentially unaltered in an electronic relationship. Although other formal requirements, such as the one that contracts must be in "writing" and "signed," can successfully be finessed through the use of carefully designed techniques.

Joe Ryan <JORYAN@suvm.acs.syr.edu> is with the Syracuse University School of Information Studies, 4-206 Center for Science and Technology, Syracuse, NY 13244-4100.

ELECTRONIC NETWORKING, Vol. 2, No. 1, Spring 1992, pp. 72–83. © 1992 Meckler

Both the Baum and Perritt and Wright books survey the essential core territory:

- a discussion of the technologies and their applications in easy to understand terminology
- examination of practical issues involving risks and controls, particularly techniques to ensure the trustworthiness and reliability of electronic messages and records
- central legal issues involving the formation of contracts, as well as evidentiary and proof issues
- the respective roles and responsibilities of network service providers and customers.

In this last area, Baum and Perritt go substantially further than Wright. They devote two major chapters to detailed analyses of and suggestions regarding the contents of trading partner agreements and third-party service provider agreements, much of which is not available elsewhere.

In addition, Baum & Perritt cover some territory not explored by Wright. Of particular interest is the concept of the EDI Clearinghouse which would be a hybrid service provider going beyond the value added networks. The clearinghouse is envisioned as an administrative, technical, and legal infrastructure providing various telecommunications and computer-based commercial trading services to facilitate electronic trade. As such it would provide an array of additional services, analogous to the check clearing services of a bank or the certification services of a post office, as well as conformance and other testing services and cryptographic key management. For his part, Wright devotes two chapters to practical electronic recordkeeping issues with particular emphasis on the implications for tax recordkeeping.

Other areas covered by Baum and Perritt deserve special mention. A comprehensive analysis of warranty and tort liability issues associated with electronic service providers is provided, as is an overview of antitrust and economic regulation. Both areas will prove invaluable to business planners in the electronic networking arena. A chapter covers federal government policies regarding access to and dissemination of information in electronic formats. Other issues including the potential use of techniques such as EDI in the procurement process as well as in rule making and adjudication are also discussed.

In sum, both books serve well as basic resources on the issues raised by electronic networking and related technologies in commerce. Baum and Perritt's work, however, has greater depth and detail of analysis and covers a number of issues not examined by Wright. On the other hand, Wright's work is probably more accessible to the layperson and to those with a greater interest in basic discussions of the covered topics. The serious "business networker," whether attorney, marketer, or service provider, will probably invest in both.

Reviewed by: Peter N. Weiss, Senior Policy Analyst and Attorney, United States. Office of Management and Budget. Office of Information and Regulatory Affairs. The views expressed herein are those of the author and do not reflect those of the agency.

Editor's Note: Also of interest are:

Emmelhainz, Margaret A. (1990). *Electronic data interchange: A total management guide.* New York: Van Nostrand Reinhold. Available: Van Nostrand Reinhold, 115 5th Ave., New York, NY 10003. Phone (800) 926-2665 Cost: $ 36.95 ISBN: 0-442-31844-8.

Payne, Judith E., & Anderson, Robert H. (1991). *Electronic data interchange (EDI): Using electronic commerce to enhance defense logistics.* Santa Monica, CA: Rand. Available: Rand, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138. Phone: (213) 393-8411. ISBN 0-8330-1124-3.

Hedlund, Patric & Meyer, Gary. (Producers). (1991). *The complete video library series of the first conference on computers, freedom & privacy.* Topanga, CA: Computers, Freedom & Privacy Video Library Project. Available: Computers, Freedom & Privacy Video Library Project; P.O. Box 912; Topanga, CA 90290; USA; E-mail: <cfpvideo@WELL.SF.CA.US> Phone: (213) 455-3915 Fax: (213) 455-1384 Format: 1/2-inch VHS videocassette, 15 videocassettes, total 19 hours, 51 minutes. Cost: $480, add $15 for shipping; in California add $39.60 sales tax; add $20 US for shipping to Canada; add $40 US for shipping to Mexico; add $60 US for shipping to Western Europe (except France), Colombia, S. Korea, Taiwan; add $71 US for shipping to Australia/NZ, Norway, Africa, Sweden, Philippines, Thailand, Saudia Arabia, Singapore; add $84 US for shipping to France, Japan, Central and South America. Individual videocassettes may be purchased at $55 each,

add $4 for shipping each in United States, in California add 8.25 percent sales tax; add $6 US for shipping to Canada; add $15 US for other international shipping.

Warren, Jim, Thorwaldson, Jay, & Koball, Bruce. (Eds.). (1991). *Proceedings of the first conference on computers, freedom & privacy*. Los Alamitos, CA: IEEE Computer Society Press. ISSN: 0-8186-2565-1. Available: IEEE Computer Society Press; 10662 Los Vaqueros Circle, P.O. Box 3014; Los Alamitos, CA 90720-1264; USA Phone: (714) 821-8380 Fax: (714) 821-4010. Format: 230 pps., 8 1/2" X 11" book. Cost: Members of IEEE-CS or Computer Professionals for Social Responsibility, $29, all others, $39, add $4 for shipping, in California add 7.75 percent sales tax.

The videocassette series provides "gavel to gavel" documentation of the First Conference on Computers, Freedom and Privacy, a forum drawing over 600 attendees and a diverse panel of speakers on March 26-28, 1991, in Burlingame, California.

This conference, frequently referred to as "The Constitutional Convention of Cyberspace," was sponsored by the Computer Professionals for Social Responsibility. Co-sponsors and cooperating organizations include: Institute of Electrical & Electronics Engineers—USA; Association for Computing Machinery; Electronic Networking Association; Electronic Frontier Foundation; Videotex Industry Association; Cato Institute; American Civil Liberties Union; ACM Special Interest Group on Software; IEEE-USA Intellectual Property Committee; ACM Special Interest Group on Computers & Society; ACM Committee on Scientific Freedom & Human Rights; IEEE-USA Committee on Communications & Information Policy; Apple Computer, Inc.; Autodesk, Inc.; Portal Communications; The WELL.

The conference hosted hundreds of people from the fields of law, computer science, law enforcement, business, public policy, government, marketing, information providing, advocacy, research, and education. Its goal was to bring together major communities and interest groups with a stake in the fundamentally new societal changes caused by information technology, and to facilitate the sharing of ideas, concerns, and experiences.

The conference featured engrossing, wide-ranging papers and panelist/audience interaction on the relationship of constitutional protection of rights and electronic access to information. The production of the videocassette series is of professional quality and captures clearly panelists' addresses, audience reactions, and formal speaker/audience question-and-answer exchanges.

Contents of the videocassette series

Tape 1
The Constitution in the Information Age
(75 mins.)

Policy proposals regarding constitutional protection, networked computers and electronic communications. Chair: Jim Warren. Speaker: Laurence H. Tribe, Professor of Constitutional Law, Harvard University Law School, "The Constitution in Cyberspace: Law & Liberty Beyond the Electronic Frontier."

Tape 2
Trends in Computers and Network
(90 mins.)

Overview and prognosis for computing capabilities and networking as they impact personal privacy, confidentiality, security, one-to-one and many-to-one communications, plus access to information about government, business, technology, and society. Chair: Peter Denning. Speakers: Peter J. Denning, Research Institute for Advanced Computer Science, "Computers Under Attack"; John S. Quarterman, Texas Internet Consulting, "The Matrix as Volksnet"; Peter G. Neumann, Computer Science Lab, SRI International, "Computers at Risk: The NRC Report and the Future"; Martin E. Hellman, Professor, Stanford University, "Cryptography and Privacy: The Human Factor"; David Chaum, Professor, Amsterdam, "Electronic Money and Beyond"; David J. Farber, Professor, Computer and Information Sciences, University of Pennsylvania, "Will the Global Village be a Police State?"

Tape 3
International Perspectives and Impacts
(75 mins.)

Other nations' models for protecting personal information and communications, and for granting access to government information, including the European Community's 1992 trans- border data flow and accountability issues; implications for privacy and personal expression. Chair: Ron Plesser. Speakers: Robert Veeder, Acting Chief, Information and Policy Branch, Office of Information Regulatory Affairs, U.S. Office of Management and Budget, Washington, DC; Tom Riley, Canadian Specialist in International Computer Privacy Issues; David H. Flaherty, Professor of History and Law, Social Science Center, University of Western Ontario, Canada; Ronald L. Plesser, Attorney, Piper and Marbury, General Counsel, U.S. Privacy Protection Study Commission.

Tape 4
Personal Information and Privacy—I
(75 mins.)

Government and private collection, sharing, marketing, verification, use, protection of, access to and responsibility for personal data, including lifestyle, work, health, school, census, voter, tax, financial and consumer information. Chair: Lance Hoffman. Speakers: Janlori Goldman, Director of Project on Privacy and Technology, American Civil Liberties Union; John Baker, Senior Vice President, Consumer and Government Affairs, EQUIFAX, Inc. Debate: Should individuals have absolute control over secondary use of their personal information? Alan F. Westin, Professor of Public Law and Government, Department of Political Science, Columbia University, New York City; Marc Rotenberg, Washington, DC Director, Computer Professionals for Social Responsibility.

Tape 5
Personal Information and Privacy—II
(75 mins.)

Ethics of "Strip Mining Data" for resale in the Information Economy. Strong international perspective. Chair: Lance Hoffman. Speakers: Simon Davies, Convenor, Faculty of Law, Privacy International, University of New South Wales, Australia; Evan Hendricks, Editor/Publisher, Privacy Times; Tom Mandel, Director, Leading Edge Values and Lifestyles Program, SRI International; Willis Ware, RAND Corporation.

Tape 6
Network Environments of the Future
(41 mins.)

Chair: Marc Rotenberg. Speaker: Eli M. Noam, Professor, School of Business, Columbia University, Center for Telecommunications and Information Studies, "Reconciling Free Speech and Freedom of Association."

Tape 7
Law Enforcement Practices and Problems
(90 mins.)

Investigation, prosecution, due process, and deterring computer crimes now and in the future; use of computers to aid law enforcement. Chair: Glenn Tenney. Speakers: Robert M. Snyder, Organized Crime Bureau, Public Safety Department, Division of Police, Columbus, Ohio; Donald Delaney, Senior Investigator, Major Case Squad, New York State Police; Dale Boll, Deputy Director, Fraud Division, United States Secret Service, Washington, DC; Don Ingraham, Assistant District Attorney, Alameda County District Attorney's Office.

Tape 8
Law Enforcement and Civil Liberties
(83 mins.)

Interaction of computer crime, law enforcement and civil liberties; issues of search, seizure, and sanctions, especially as applied to networked information, software, and equipment. Chair: Dorothy Denning. Speakers: Sheldon T. Zenner, Attorney, Katten, Muchin, and Davis, Chicago; Kenneth Rosenblatt, Deputy District Attorney, Santa Clara County District Attorney's Office; Mitchell Kapor, President, Electronic Frontier Foundation; Mike Gibbons, Supervisory Special Agent, Federal Bureau of Investigation; Cliff Figallo, Executive Director, The WELL; Sharon Beckman, Attorney, Silverglate and Good, Boston; Mark Rasch, Trial Attorney, U.S. Department of Justice.

Tape 9
Legislation and Regulation
(82 mins.)

Legislative and regulatory roles in protecting privacy and insuring access; legal problems posed by computing and computer networks; approaches to improving government processes; limits on legislation. Chair: Bob Jacobson. Speakers: Craig Schiffries, Congressional Science Fellow, Subcommittee on Technology and the Law, Senate Judiciary Committee; Bill Julian, Chief Counsel, Utilities and Commerce Committee, California State Assembly; Jerry Berman, Director, Information Technology Project, American Civil Liberties Union; Paul Bernstein, Attorney, LawMUG BBS and Electronic Bar Association Legal Information Network; Elliot T. Maxwell, Assistant Vice President for Corporate Strategy, Pacific Telesis; Steve McLellan, Policy Strategist, Washington Utilities and Transportation Commission, Olympia.

Tape 10
Computer-Based Surveillance of Individuals
(90 mins.)

Monitoring of electronic mail, public and private teleconferences, electronic bulletin boards, electronic "publications" and their subscribers; computer-aided monitoring of individuals, work performance, buying habits and personal lifestyles. Chair: Susan Nycum. Speakers: Judith F. Krug, Director, Office for Intellectual Freedom, American Library Association; Karen Nussbaum, Executive Director, 9 to 5 National Association of Working Women; Gary T. Marx, Professor of Sociology, Massachusetts Institute of Technology; David H. Flaherty, Professor of History and Law, Social Science Center, University of Western Ontario, Canada.

Tape 11
Security Capabilities, Privacy and Integrity
(69 mins.)

Chair: Dorothy Denning. Speaker: William A. Bayse, Assistant Director, Technical Services, Federal Bureau of Investigation, Washington, DC, "NCIC—2000: Balancing Computer Security Capabilities with Privacy and Integrity."

Tape 12
Electronic Speech, Press and Assembly
(91 mins.)

Freedoms of electronic speech, public and private electronic assembly, and electronic publishing; issues of prior restraint and chilling effects of monitoring on freedoms; possible justifications for monitoring; alternatives. Chair: Eric Lieberman. Speakers: Lance Rose, Attorney, Wallace & Rose, New York City; Jack Rickard, Editor, BOARDWATCH MAGAZINE, Boardwatch Online Information Service; George Perry, Vice President and General Counsel, Prodigy Services Co.; John McMullen, Consultant and Journalist, Newsbytes, and McMullen & McMullen, Inc.; Eric Lieberman, Attorney, Rabinowitz, Baudin, Standard, Krinsky & Lieberman, New York City; David Hughes, Electronic Citizen and General Partner, Old Colorado City Communications.

Tape 13
Access to Government Information
(89 mins.)

Implementing individual and corporate access to federal, state, and local information about communities, corporations, legislation, administration, the courts and public figures; allowing access while protecting privacy. Chair: Harry Hammitt. Speakers: Harry Hammitt, Editor and Publisher, ACCESS REPORTS, Inc.; Katherine F. Mawdsley, Associate University Librarian, University of California at Davis; David Bright Burnham, Co-Director and Writer, Transactional Records Access Clearinghouse; Robert Veeder, Acting Chief, Information Policy Branch, Office of Information Regulatory Affairs, U.S. Office of Management and Budget, Washington, DC.

Tape 14
Ethics and Education
(83 mins.)

Ethical principles for individuals, system administrators, organizations, corporations and government; copying of data, copying of software, distributing confidential information; relations to computer education and computer law. Chair: Terry Winograd. Speakers: Dorothy Denning, Systems Research Center, Digital Equipment Corporation; Donn B. Parker, Senior Management Consultant, SRI International; Richard Hollinger, Associate Professor, Department of Sociology, University of Florida; John Gilmore, Generalist, Cygnus Support; Jonathan Budd, Program Manager, Law Enforcement Computer Crime, National Institute of Justice; Sally Bowman, Director, Computer Learning Foundation.

Tape 15
Where Do We Go from Here?
(83 mins.)

Perspectives, recommendations, and commitments of participants from differing interest groups, proposing next steps they will pursue to protect personal privacy, protect fundamental freedoms, and encourage responsible private-sector and public-sector policies and legislation. Chair: Jim Warren. Speakers: Paul Bernstein (see Tape 9); Mary J. Culnan, Associate Professor, School of Business Administration, Georgetown University; David Hughes (see Tape 12); Don Ingraham (see Tape 7); Mitchell Kapor (see Tape 8); Eric Lieberman (see Tape 12); Donn B. Parker (see Tape 14); Craig Schiffries (see Tape 9); Robert Veeder (see Tape 3).

The printed proceedings make an excellent supplement to the videotape series, acting as a comprehensive, edited transcript. Editing of papers was done for clarity only, and the proceedings are an accurate reflection of the conference's contents. A helpful index is included.

The Second Conference on Computers, Freedom & Privacy is scheduled for March 18-20, 1992, L'Enfant Plaza Hotel, Washington, DC. Information may be obtained from Professor Lance Hoffman, Department of Electrical Engineering and Computer Science, George Washington University, Washington, DC 20052; (202) 994-4955.

Reviewed by: Bruce Flanders <Flanders@ukanvm.bitnet> Director of Technology, Kansas State Library, Topeka, KS.

Egan, Bruce L. (1991). *Information superhighways: The economics of advanced public communication networks*. Norwood, MA: Artech House. 187 pp. Available: Artech House, 685 Canton St., Norwood MA 02062. Phone: (800) 225-9977 (617) 769-9750 Cost: Hard: $55 ISBN: 0-89006-474-1.

This is a hedgehog of a book, which reveals its treasures slowly and stubbornly. But the treasures are real, whether or not you share Egan's approach or conclusions. *Information Superhighways* will force you to rethink your assumptions about the relationship between the evident public good of ubiquitous

broadband communications and the economic structures required to construct and sustain it.

Before going on, I need here to state an interest: this book is part of the Artech House Telecommunications Library, a series edited by my colleague Dr. Vinton G. Cerf. I do not believe, however, that this relationship has any material effect on my comments.

In *Information Superhighways*, Egan provides a broad-brush survey and analysis of the economics and related issues involved in the construction and dissemination of high speed public information networks, termed Universal Broadband Networks (UBNs). The book briefly reviews the relevant technologies but focuses primarily on the economic structures that would underlie their widespread dissemination.

There are important discussions of the regulatory framework in which the principal stakeholders operate, and of the capital and accounting structures that are expected to form the foundation for the Bell Operating Companies (BOCs), cable and satellite company, and network decisions about implementing UBNs. There are also less comprehensive analyses of the public policy implications of UBN implementation, and of the related but distinct political (small and large p) framework for both public and private decision making in this arena. Egan concludes, somewhat wistfully (p. 178):

> In summary, it is important that a public consensus be reached on the proper goals of communication policy, for the status quo alternative may be neither privately nor publicly efficient. In fact, it may further exacerbate the technology-policy crisis. UBNs may never have a chance if the political tug-of-war between the free-marketers and infrastructure faction leaves the country with no public policy.

In Egan's view, this tug-of-war is likely to determine whether or not UBNs are implemented. The core issues, from his perspective, are regulatory and financial. As he points out (p. 2), "current public policy is inconsistent with a paradigm for sharing and interconnection. Regulatory and legal policies encourage structural separation of networks through a host of asymmetric rules across industry segments." Later on, he adds:

> If policy-makers wish to further the infrastructure approach to public communications networks, they should remove impediments to telephone companies and cable companies, two important infrastructure players, by proposing even higher depreciation rates, removal of business restrictions, and the like.

Egan's analysis of the regulatory and accounting thickets confronted by telephone and cable companies, in chapters 4 and 8 in particular, is itself worth the price of admission. *Information Superhighways* is a solid and substantial first look at an exceedingly important, controversial, and complex topic, one that goes to the core of the hopes and dreams for the National Research and Education Network (NREN). What else might have been desired?

First, the topic does not need to be quite so inaccessible. There are far too many acronyms. Do we really need POTS for plain old telephone service? Many of these acronyms raise their head only occasionally above the dense prose, never to be seen again. At times this work reads like a Russian novel, without the characters listed in the back.

Some important elements of the economics of networks are, in my view, insufficiently discussed. For example, Egan is very creative about elaborating (from limited data) the potential costs of UBNs. But on the crucial demand side, he merely says (p. 146):

> There is precious little credible research quantifying the added value of broadband technology to society, and even less evidence of what people (or their government representatives) are actually willing to pay for it.

True. But there are pointers toward some important alternative scenarios. Egan does not explore them. Second, the international implications of UBNs, the essential international medium, are only briefly analyzed. Furthermore, Egan overemphasizes the importance of network companies saying (p. 151): "ultimately, it is communication network suppliers who will determine the direction of the development and deployment of UBNs, or whether they occur at all." At the same time he nearly ignores the crucial political decisions that will, in large measure, determine whether or not UBNs will happen.

Will candidates try to ride the ubiquitous electronic communications wave as an issue to prove that they've got a handle on the "vision thing?" Will a perceived consumer demand create a consensus about the need for UBNs among states, utility commissions, local government overseers of cable companies, the FCC, and private citizen groups? Will the current despair over public K-12 education generate agreement about the need for a computer, bearing important learning materials, in every home and classroom? As a result of an evolving national consensus about the public good, will competing stakeholders, like the BOCs, cable companies, satellite providers, and broadcasters, be encouraged to cooperate to build, stock, and disseminate national communication networks?

The national response to these questions, painfully constructed through an ongoing debate about the relative value of this and other infrastructure investments, is more likely to determine the course of infrastructure construction and dissemination than the network supplier cost and accounting decisions that Egan so carefully describes.

But these limitations do not diminish the value of *Information Superhighways*. Instead, they point toward future dialogue, experimentation, and research. *Information Superhighways* is an essential primer, delineating many of the basic economic, regulatory, and procedural issues involved in the development and deployment of broadband communication networks. The fundamental issues of value and public policy remain and urgently require continuing discussion.

By definition, in our democracy, these issues can never be resolved. But *Information Superhighways* can, at least, help us to understand how to talk about them.

Reviewed by: John R. Garrett, Ph.D. <JGarrett/cnri@cnri@mcimail.com> Information Resources, Corporation for National Research Initiatives (CNRI), 1895 Preston White Drive, Suite 100, Reston, VA 22091. Phone 617-631-3419. Fax 617-631-4395

Neubauer, Karl Wilhem, and Dyer, Esther, R. (Eds.) (1990). *European library networks*. Norwood, NJ: Ablex. 435 pp. Available: Ablex Publishing Company, 355 Chestnut St., Norwood, NJ 07648. Phone: (201) 767-8450. Cost: $75. ISBN: 0-89391-157-7.

The problem with books on any aspect of computer technology is that they can so easily be overtaken by fast developing events. This book is no exception. Neubauer and Dyer offer a compilation of articles on some of the European computerized library networks describing the situation at the end of 1989. Already this source is somewhat out of date.

This work covers the wide range of library networks in Europe, demonstrating how the use of networking inevitably reflects the cultural and administrative structures of each country. European library networks include the centralized structure of France, where a national policy for university library networks and major higher education institutions was created through DBMIST. The federal structure of Germany has led to the establishment of a variety of library networks that meet the local needs. Care is being taken to allow for potential integration with systems in the other federal states. The Project on Integrated Catalogue Automation (PICA) in the Netherlands is funded by the Dutch government. The aim there is to establish an automated library network on a national scale.

In the United Kingdom a number of library networks have developed in an uncoordinated manner, with relatively little centralized preplanning and direction. Some would say that this is the story of Britain in the 1980s. As the country now slips to the bottom of every economic league, it awaits rescue by the "market," with the government ideologically prevented from trying other solutions. It was unfortunate for British libraries that the technology matured in the 1980s. This encouraged central planning and investment in library networks. But this approach was adopted at a time when the prevailing political culture was opposed to intervention and policy.

In addition to the major descriptions of French, German and British library networks, this source provides brief overviews of particular developments in Austria, Denmark, Italy, Norway, Sweden, and Switzerland. In addition, there is a useful overview of European library networks, their various structures, and the type of services available.

There is an extensive bibliography on library networks worldwide, organized into sections on applications, on countries and continents, and on individual library networks. The presentation of the bibliography, with no introduction to indicate its purpose or scope, is indicative of one of the failings of the book; a lack of overall editing. Most of the contributions need to be set into their national contexts, although there is a general overview of

developments in automated bibliographic networks in the United Kingdom.

Many of the authors will be well known to readers of this journal, but a short biography of each contributor would have been useful. All the contributions are provided by the managers or directors of the particular networks. As a result, many of the networks are merely described without a critical perspective. One of the most interesting and refreshing contributions, however, is by Bernard Gallivan on SCOLCAP, the Scottish library network.

SCOLCAP began in 1973 with the aim of assisting libraries in their tasks of acquiring and cataloging books. SCOLCAP fell apart by the late 1980s for a number of reasons. A principal reason was the sudden availability of affordable integrated library systems, which enabled libraries to have their own systems, under their own control. What makes Gallivan's contribution of greater interest is his general conclusions on library cooperation and networks drawn from the failure of SCOLCAP. He is skeptical of libraries cooperating in any meaningful way, because (p. 269) "the reality is that librarians actually prefer being on their own, doing their own thing....Librarians actually prefer being masters of their own empires, no matter how large or small."

This book is a useful snapshot of the state of some of the European networks at the end of 1989. But given the fast pace of change, periodical articles are probably the best way to keep up to date in this fast developing area.

Reviewed by: Michael Breaks <LIBMLB@CLUST.HW.AC.UK>, ENRAP European Editor, Library, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, Scotland. Phone: + 44 31 449 5111 Fax: + 44 31 451 3164.

McClure, Charles R., Bishop, Ann, Doty, Philip, & Rosenbaum, Howard. (1991). *The National Research and Education Network (NREN): Research and policy perspectives*. Norwood, NJ: Ablex. 744 pps. Available: Ablex Publishing, 355 Chestnut St., Norwood, NJ 07648. Phone: (201) 767-8450.

The authors have made a significant contribution to the current debate about the shape of the National Research and Education Network (NREN). The major contributions of the work are to provide the reader with:

- information and background on the NREN, including major source documents
- reviews of a number of research efforts on issues related to national networking
- a review of relevant literature and an extensive bibliography
- an outline of the major unresolved policy issues along with recommendations for action.

The resource book is organized into nine chapters and fourteen appendices. The first 170 pages of text represent the original contributions of the authors followed by an extensive bibliography. Chapters 5, 6, and 7 originally appeared as background papers prepared for the U.S. Congress, Office of Technology Assessment, and portions of Chapters 8 and 9 were part of those background papers as well.

The fourteen appendices are, for the most part, government documents, including the entire text of four versions of the Senate bill prior to its being signed into law. President Bush signed the High Performance Computing Act (P.L. 102-194) into law on December 9, 1991, a few months after the book was published. Thus, the final version of the enabling legislation is not included. Students and scholars will find it necessary to consult this final version of the joint House and Senate bill to close the circle on the evolution of NREN legislation. [Editor's Note: See the editorial in this issue of ENRAP].

Other appendices represent background papers prepared for the Senate Committee on Commerce, Science and Transportation, the Senate Committee on Energy and Natural Resources, the President's Office of Science and Technology, the Office of Technology Assessment, and the Congressional Research Service. It is useful to have all the resource materials brought together in one reader.

The book has both an author index and a comprehensive subject index. A glossary is provided to help the reader uncover the difference between CREN and CAUSE or CNI and CNRI, as well as to provide useful definitions for "technospeak."

The two most interesting chapters for this reviewer were Chapters 3 and 9, which by themselves made the book worthwhile reading. Chapter 3 reviews the promised benefits expected to emerge, the potential problems likely to arise, and the major policy issues involved in the debate over the NREN. The chapter draws from the literature to present

three very useful charts of benefits, problems, and policy issues. The discussion of each element listed in the charts is clear and well written. A very abbreviated summary of each of the tables follows. The benefits identified include:

- enhancing U. S. competitiveness
- access to supercomputers
- removing constraints from the research process
- increasing collaboration among researchers
- increasing the rate of knowledge and technology transfer.

The problems with the existing network are seen to include:

- fragmentation
- limited capacity of existing networks
- lack of user friendliness.

The problems that may arise with the development of the NREN include:

- adverse social impacts of the NREN on science
- increasing the burden on scientists
- threats to security and privacy
- technological problems

The key policy issues include:

- design and construction
- size of the NREN
- access
- equity and fairness
- transition from existing networks to the NREN
- management structure
- maintenance and operation
- legal issues
- finances and cost recovery
- transition to the private sector
- network use
- user education, training, and support
- security and privacy
- censorship

This discussion of the policy issues will be especially helpful to interest groups, such as public libraries, academic, and research institutions, for developing position papers and designing lobbying efforts to influence the shape of the "five-year plan" specified in the legislation.

The policy issues identified in chapter 3 are expanded and recast as important unanswered questions in chapter 9. The issues, as presented in chapter 9, reflect a scholar's look at questions to be addressed in future research. This is a useful tool to encourage important <network> research. The value of the questions for the policy analyst or interest group leader is to make these individuals aware of the vast unknown. The hope is that policies that are developed will make allowances for evolution as these questions are answered. It seems to this reviewer that the most critical issue areas requiring immediate attention include education, training, and support; understanding the role of information technology in research work; access; and network design and management.

The authors present nine recommendations, in the order of importance:

- Conduct survey of existing network users/ policies
- Design the NREN in light of user information needs and behavior
- Require direct support for network training
- Obtain greater involvement from the library community
- Provide better documentation and directories
- Establish a lead federal agency for NREN development
- Plan for the management of the NREN

- Develop mechanisms to improve communication between network engineers/managers and network users
- Conduct additional research.

In my view, in light of the passage of the NREN enabling legislation, the establishment of a lead federal agency seems most critical. Such an agency should have broad representation of the user communities in academic and research institutions, and the participation of public libraries becomes a high priority.

Chapters 5, 6, and 7 will be of particular interest to researchers and science librarians trying to understand and cope with the impact of electronic networks on the work of professionals in the scientific and technical community. In chapter 5, Susan Koch does an excellent job of reviewing and synthesizing the literature on "Electronic Networks and Science." Chapter 6 presents the results from an empirical study conducted by the authors examining the impact of networks on research. Chapter 7 is an interesting scholarly examination of the social norms of science that the authors relate to comments made by researchers who participated in the empirical study on the use of networks to improve research quality.

The likely audiences for the book will include information scientists, librarians, academic administrators, students, information policy analysts, network architects, and various interest groups. The academic and research community will find more of interest than will those interested in "public access computing for the masses." The authors focus on the "user's perspective," but the users they describe are primarily scientific researchers. Broadening the perspective would have strengthened the book for policy analysts concerned with developing policies to meet the needs of a more inclusive user community. The authors acknowledge the need to expand the user community to include the public at large, but not as much attention is paid to those users. The greatest value of the book is as a resource for information policy analysts and academic administrators in their efforts to develop a better understanding of the broad social policy implications of the NREN.

Reviewed by: Carolyn M. Gray <GRAY@BINAH.CC.BRANDEIS.EDU>

Associate Director, Brandeis University Libraries, Box 9110,

Waltham, MA 02254-9110. Phone: (617) 736-4700 Fax: (617) 736-4675. Carolyn is past President of the Library and Information Technology Association (LITA).

Resource Reviews critically examines current information resources on, or about, electronic networks. Of particular interest are information resources that:

- Report recent research related to networks
- describe and evaluate network applications and use in a variety of settings
- Discuss network standards, management, regulation, and governance
- Explore network policy issues and the impact of electronic networks on individuals, groups, or society from a variety of disciplinary perspectives
- Promote the successful use of electronic networks
- Provide innovative views and approaches to electronic networking.

The Resource Review Editor encourages publishers of material of potential interest to our readers to send these items to the Resource Review Editor. Reviewers interested in writing for *Electronic Networking* are also encouraged to contact the Resource Review Editor.

Items Received

Aggarwal, V. (1991, December). *Mid-level networks potential technical services*. (RFC 1991). Princeton, NJ: JvNCnet Computer Network. Available: FTP: FTP.NISC.SRI.COM Directory: INTERNET-DRAFTS File: RFC1990.TXT E-mail: SENDRFC@JVNC.NET Message: RFC 1991

Antell, Karen <antell@mcnc,org>. (Ed.). (1991 and ongoing). Communications for North Carolina Education, Research and Technology (CONCERT) *Notes*. Research Triangle Park: MCNC Center for Communications. Available: MCNC Center for Communication,P.O. Box 12889, Research Triangle Park, NC 27709-2889. Phone: (919) 248-1863.

Cerf, Vinton G. <isoc@nri.reston.va.us>. *Internet Society Publications*. Reston, VA: Internet Society. Available: FTP: NNSC.NSF.NET Directory: ISOC File: INDEX-ISOC

Clark, D., Chapin, L., Cerf, V., Braden, R. & Hobby, R. (1991, December). *Towards the future Internet architecture*. (RFC 1287). Available: FTP: FTP.NISC.SRI.COM Directory: INTERNET-DRAFTS File: RFC1287.TXT E-mail: SENDRFC@JVNC.NET Message: RFC 1287

Computer Systems Policy Project. (1991). *Expanding the vision of high performance computing and communications: Linking America for the future*. Washington, DC: Computer Systems Policy Project. 14 pages and accompanying video tape. Available: Computer Systems Policy Project, 1735 New York Ave., NW., Suite 500, Washington DC 20006. Cost: $20.

Cook, Gordon. (1992). *The National Research and Education Network: Whom shall it serve*? Ewing, NJ: Cook Network Consultants. 35 pp., mimeograph. Available: Cook Network Consultants, 431 Greenway Ave., Ewing, NJ 08618. Cost: $25.

Government Accounting Office. (1991, November 20). *Computer security: Hackers penetrate DOD computer systems*. (GAO/IMTEC-92-5) Washington, DC: Government Accounting Office. Available: FTP: CU.NIH.GOV Directory: GAO-REPORTS File: IMTEC-92.T5 For further information, contact Jack L. Brock, Jr. <KH3@CU.NIH.GOV>.

Government Accounting Office. (1991, September). *High-performance computing: High-speed computer networks in the United States, Europe, and Japan*. (GAO/IMTEC-91-69) Washington, DC: Government Accounting Office. Available: FTP: CU.NIH.GOV Directory: GAO-REPORTS File: IMTEC-91.R69 For further information, contact Jack L Brock, Jr. <KH3@CU.NIH.GOV>.

Government Accounting Office. (1991, July). *High-performance computing: Industry uses of supercomputers and high-speed networks*. (GAO/IMTEC-91-58) Washington, DC: Government Accounting Office. Available: FTP: CU.NIH.GOV Directory: GAO-REPORTS File: IMTEC-91.R58 For further information, contact Jack L Brock, Jr. <KH3@CU.NIH.GOV>.

Government Printing Office. (1992, January). *GPO/2001: Vision for a new millennium*. Washington, DC: Government Printing Office. Available: GPO, North Capital & H Streets NW, Washington DC 20407. Phone: (202) 275-2051.

High-Performance Computer Act of 1991., P.L. 102-194. (1991, December 9) Available: FTP: NIS.NSF.NET <35.1.1.48> Directory: NSFNET File: NRENBILL.TXT or E-mail: NIS-INFO@NIS.NSF.NET Message: SEND NRENBILL.TXT

Lottor, M.<mkl@nisc.sri.com>. (1992, January). *Internet growth (1981-1991)*. (RFC 1296). Menlo Park, CA: SRI International. Available: FTP: NIC.NEAR.NET Directory: DOCS File: INTERNET-GROWTH

Martin, J. (1990, December). *There's gold in them thar networks! Or searching for treasure in all the wrong places*. (RFC 1990). Columbus, OH: Ohio State University. Available: FTP: FTP.NISC.SRI.COM Directory: INTERNET-DRAFTS File: RFC1990.TXT E-mail: SENDRFC@JVNC.NET Message: RFC 1990

National Science Foundation. (1992). *NSFNET acceptable use policy*. Washington, DC: National Science Foundation. Available: FTP: NIS.NSF.NET Directory: NSFNET File: NETUSE.TXT or NETUSE.PS (postscript version) or E-mail: NIS-INDO@NIS.NSF.NET Message: SEND NETUSE.TXT

National Science Foundation. (1992). *Project Development Plan for the continuation and enhancement of NSFNET backbone services*. Washington, DC: National Science Foundation. Available: FTP: NIS.NSF.NET Directory: NSFNET File: DEVPLAN.TXT or FTP: NIS.NSF.NET Directory: CISE File: PDP.PS (postscript version).

NEARNET. (1991, December 12). *NEARNET technical user seminar*. Cambridge, MA: Bolt, Beranek

& Newman, Inc. Available: Bolt, Beranek & Newman, Inc., 10 Moulton St., Cambridge, MA 02138. E-mail: NEARNET-STAFF@NIC.NEAR.NET

North American Directory Forum. (1992, January). *User bill of rights for entries and listings in the public directory.* (RFC 1295). Network Working Group. Available: FTP: NNSC.NSF.NET <128.89.1.178> Directory: RFC File: RFC1295.TXT

NorthWestNet, Inc. (1992). *NorthWestNet User Services Internet Resource Guide.* Bellevue, WA: NorthWestNet. Available: NorthWestNet, 15400 SE 30th Place, Suite 202, Bellevue, WA 98007. Phone: (206) 562-3000 Fax: (206) 562-4822.

Office of Science and Technology Policy. Federal Coordinating Council for Science, Engineering, and Technology. Committee on Physical, Mathematical, and Engineering Sciences. (1992). *Grand challenges 1993: High performance computing and communications, the FY 1993 U.S. research and development program* [to supplement the President's Fiscal Year 1993 Budget]. Washington, DC: National Science Foundation. 65 pp. Available: presently free from NSF, 1800 G Street, N.W. Washington, DC 20550. Phone: (202) 357-7861 or via NTIS Cost: $19 + $3 shipping.

Proceedings for wide area information server workshop. (1992, February 3-4). Brett, George (Conference Chairperson), Kahle, Brewster, & Oberlin John (Conference Co-chairperson). Research Triangle, NC: MCNC Center for Communications. Available: MCNC Center for Communications, P.O. Box 12889, 3021 Cornwallis Road, Research Triangle Park, NC 27709-2889. Phone: (919) 248-1800.

SRI International <NISC@NISC.SRI.COM> (1992). *TCP/IP CD-ROM.* Menlo Park, CA: SRI International. Available: SRI International, Network Information System Center, Room EJ291, 333 Ravenswood Ave. Menlo Park, CA 94025 Phone: (415) 859-NETS or 859-3695 Fax: (415) 859-6028 Cost: $195.

Telematics and informatics. Singh, Indu B. (Editor-in-Chief). (Quarterly, ongoing). Elmsford, NY: Pergamon Press. Available: Pergamon Press, Inc., Front & Brown Sts., Riverside, NJ 08075-1197. Phone: (800) 257-5755 or (609) 461-6500.

Zimmerman, D. (1991, December). *The finger user information protocol.* (RFC 1288). Available: FTP: FTP.NISC.SRI.COM Directory: INTERNET-DRAFTS File: RFC1288.TXT E-mail: SENDRFC@JVNC.NET Message: RFC 1288

Contributor Guidelines

General Considerations

Electronic Networking: Research, Applications, and Policy (ENRAP) is a cross-disciplinary journal that provides coverage of the rapidly growing use of telecommunications networks to provide information services and products. The purpose of the journal is to describe, evaluate, and foster understanding of the role and applications of electronic networks. Moreover, *ENRAP* intends to promote and encourage the successful use of electronic networks. The journal is of interest to network users, managers, and policymakers in the academic, computer, communication, library, and government communities. Additional information about the journal, its purpose, scope, and production schedule, can be found in "From the Editor," in Volume 1 , Number 1.

The journal publishes papers and essays that describe research findings related to electronic networks, examine current and potential applications of electronic networking, and identify and assess policy issues related to networking. It also features reviews and announcements of print and electronic resources of interest. Initially, the journal will be issued in print format, but, eventually, the editors hope to publish it in electronic format.

Manuscripts

Send manuscripts (three copies) to the Editor: Dr. Charles R. McClure, School of Information Studies, 4–206 Center for Science and Technology, Syracuse University, Syracuse, NY 13244–4100. Authors will receive an acknowledgment when the editor receives the manuscript. Authors may wish to discuss possible contributions with the editors prior to submitting completed papers.

In order to expedite review and publication, authors must also submit a diskette containing the final draft of their papers. Authors are asked to submit 3.5-inch or 5.25-inch floppies with the contribution in any of the following formats: WordPerfect (for IBM), Microsoft Word (for IBM or Macintosh), MacWrite, or ASCII. Indicate plainly on the disk both the name and format of the file. If your paper includes notes, do *not* use the footnote/endnote function in your word processor. Simply type the notes at the end. If the authors cannot provide documents in these formats, they should contact the Meckler Corporation (Doreen Beauregard, 203–226–6967).

All manuscripts should be submitted to the editor in triplicate, double-spaced on one side of good quality 8.5 by 11-inch white paper. There should be margins of 1-inch on the top, bottom, and sides, and each page of the manuscript should be numbered. The manuscript should consist of several parts:

- *Title page*: With each author's name, address, affiliation, and e-mail address.
- *Abstract page*: One paragraph of 200 words or less that gives a descriptive summary of the paper, emphasizing its key points and conclusions.
- *Biographical Statement Page*: A 50- to 75-word biographical statement with information about each author's current position, mailing address, educational background, and other information as appropriate.

■ *Text*: The paper must meet the usual standards of organization, clarity, syntax, spelling, and the like. All submissions are subject to critical review, and the editors reserve the right to make the final determination of any manuscript's suitability for publication. In order to maintain confidentiality, there should be no indication of authorship other than on the title page.

■ *Notes*: Explanatory only and cumulated immediately after the main body of the paper.

■ *Reference list*: In APA style; see the following section on References and Notes.

■ *Length*: Typical research papers are ordinarily between 20 and 25 double-spaced pages long, while essays and reports of applications are between 15 and 20 double-spaced pages long. The editors should be contacted with any questions about appropriate length of contributions.

References and Notes

All citations and references shall be in APA (American Psychological Association) style. References are cited in the text by putting the last name of the author and the date in parentheses, for example (Smith, 1989). If the author's name is used in the text, then only the date should be in parentheses. Author names in the reference list should include first name and middle initial.

The reference list should appear at the end of the text and include only works cited in the text. The list should be arranged alphabetically by the authors' last names. Multiple entries for the same author should be arranged chronologically, with the earliest date listed first. Two or more publications by the same author should be listed in both the text and reference list by letters, for example (Smith, 1989a). Authors should consult the third edition of the APA style manual for further instructions.

Notes should be numbered consecutively and appear at the end of the text.

Tables and Figures

Each table and figure should be submitted in camera-ready form on a separate sheet of paper. All tables and figures should be mentioned in the text and numbered consecutively. Numbers and titles, in mixed upper and lower case, should be printed in pencil on the back of each table and figure, along with the brief title of the paper and the appropriate page number. Tables should also be submitted electronically just as the authors wish them to appear in the journal.

The editors encourage contributors to include figures, illustrations, tables, photographs, and other types of graphics as a means of increasing the clarity and readability of the paper.

Review Process

All submitted papers are reviewed by members of the Editorial Board and other experts as appropriate. Review is done expeditiously. Criteria considered when reviewing papers include, but are not limited to

■ Importance of the topic addressed
■ Originality of the author's treatment of the topic
■ Clarity and organization of content
■ Presentation of information, especially the use of tables and figures
■ Appropriateness of the problem, research design, and methodology (if the paper is a research study)
■ Use of and reference to appropriate literature
■ Potential interest to readers.

The editors reserve the right to make the final determination of any contribution's suitability for inclusion.

Resource Reviews

General Guidelines for Reviewers: The reviews should represent the resource fully and fairly, balancing the author's aims and results. The review should say more about the source than the reviewer. Avoid personalization of issues or people. We encourage comments on special features of the resource. For books, characteristics such as layout, illustrations, pricing, typeface, and indexing should also be considered.

The journal includes reviews of paper and electronic sources of potential interest to readers. Reviews should be descriptive (what is the source about?) and evaluative (why is/isn't the source worth the user's time or expense?). Would the user be better served using some other resource?

■ *Publishers*: Publishers and producers are invited to send books, reports, and electronic items to the Resource Editor for potential review.

■ *Reviewers*: The Resource Editor invites readers to volunteer their talents as reviewers. Write to the Resource Editor stating your interests, area of expertise, or competence.

■ *Deadlines*: Review deadlines are as follows throughout the year: January 1, April 1, July 1, and October 1.

■ *Review Length and Type*: 500 to 1,000 words for a short, critical review of a single work; 1,000 to 1,500 words for an in-depth appraisal or evaluation essay of more than one resource.

■ *Review Citation Format*: Follow APA style where appropriate. Consult Resource Editor for citation format for distinctive electronic resources.

Sample Book:

> McClure, Charles R., Bishop, Ann P., Doty, Philip, & Rosenbaum, Howard. (1991). *The National Research and Education Network (NREN): Research and policy perspectives.* Norwood, NJ: Ablex Press.
>
> Reviewed by: Joe Ryan, School of Information Studies, Syracuse University, Syracuse, NY 13244.

Sample Report:

> Gould, Stephen. (1990). *The federal research internet and the National Research and Education Network: Prospects for the 1990s.* Washington, D.C.: Library of Congress, Congressional Research Service.
>
> Reviewed by: Joe Ryan, School of Information Studies, Syracuse University, Syracuse, NY 13244.

More detailed explanation of guidelines and suggestions for resource reviewers can be found in "Resource Review Guidelines," available from the Resource Review Editor.

Additional Information

Additional information or questions regarding guidelines for contributors should be forwarded to the editors: Charles R. McClure, School of Information Studies, Syracuse University, Syracuse, NY 13244 (CMCCLURE@SUVM.ACS.SYR.EDU); Ann Bishop, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 426 David Kinley Hall, 1407 W. Gregory Drive, Urbana, IL 61801 (abishop@uiuc.edu); Philip Doty, Graduate School of Library and Information Science, University of Texas at Austin, Austin, TX 78712-1276 (pdoty@utxvm.cc.utexas.edu); or Joe Ryan, School of Information Studies, Syracuse University, Syracuse, NY 13244 (joryan@suvm.acs.syr.edu).

Electronic Networking

at

Computers in Libraries '93

Call for Papers
&
Convener Guidelines

"Electronic Networking" a conference within a conference of Computers in Libraries '93 is scheduled for February 28 - March 3, 1993 in the Sheraton Washington Hotel, Washington, DC.

Short Course (preconference workshops) will be held on Sunday, February 28. Three full days of programming will begin on Monday, March 1, 1993.

Prospective speakers are invited to propose a 40 minute presentation for the conference. Conveners may plan three sessions, four sessions, a full day of seven sessions or more. Conveners may also propose Short Course presentations.

Those with interest in serving as a convener should be in touch with **Nancy Nelson** at (203) 226-6967 (Meckler, 11 Ferry Lane West, Westport, CT 06880).

Conveners must provide a list of speakers and their topics (as fully completed as is possible) by **August 15, 1992.**

Final program details are due no later than **September 15, 1992.**

Executive summaries from all presenters are due **January 1, 1993.**